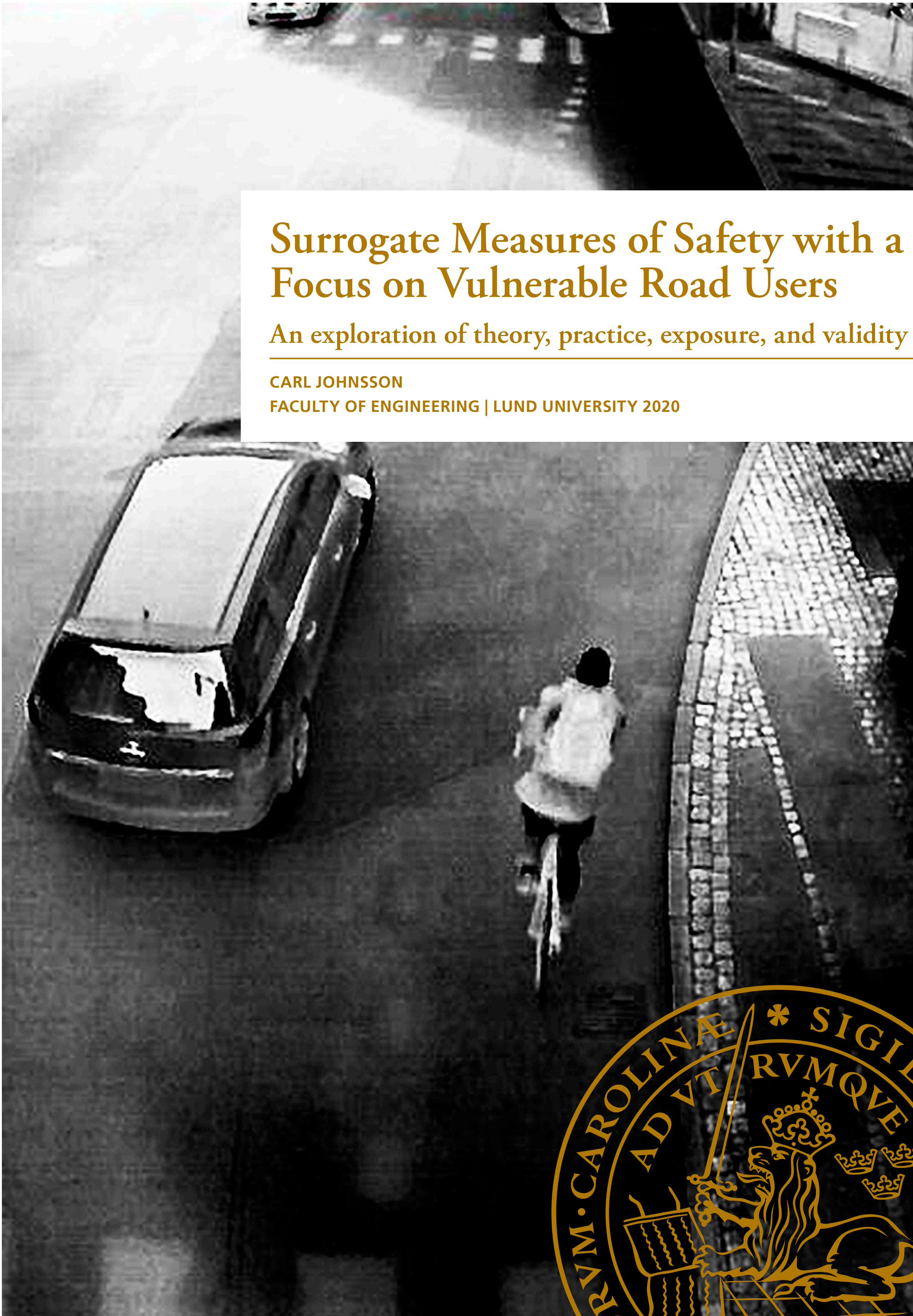
Surrogate Measures of Safety with a Focus on Vulnerable Road Users

An exploration of theory, practice, exposure, and validity

Johnsson, Carl

2020

Document Version:
Publisher's PDF, also known as Version of record



Surrogate Measures of Safety with a Focus on Vulnerable Road Users

An exploration of theory, practice, exposure, and validity

CARL JOHNSSON
FACULTY OF ENGINEERING | LUND UNIVERSITY 2020

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

Lund University
Faculty of Engineering
Department of Technology and Society
Bulletin 320
ISBN 978-91-7895-648-7
ISSN 1653-1930
CODEN: LUTVDG/(TVTT-1061)1-101

Surrogate Measures of Safety with a Focus on Vulnerable Road Users

An exploration of theory, practice, exposure, and validity

Carl Johnsson

DOCTORAL DISSERTATION

by due permission of the Faculty of Engineering, Lund University, Sweden.
To be defended at the Faculty of Engineering, John Ericssons väg 1, in auditorium V:C on the 30th of October at 15:00

Faculty opponent
Professor Tarek Sayed, University of British Columbia, Canada

Organization LUND UNIVERSITY	**Document name** Doctoral dissertation
	Date of issue 2020-10-30
Carl Johnsson	**Sponsoring organization**

Title and subtitle

Surrogate Measures of Safety with a Focus on Vulnerable Road Users: An exploration of theory, practice, exposure, and validity

Abstract

Surrogate measures of safety (SMoS) are meant to function as tools to investigate traffic safety. The term *surrogate* indicates that these measures do not rely on crash data; instead, they focus on identifying safety critical events (or near-crashes) in traffic, which can be used as an alternative to crash records.

The overall aim of this thesis is to explore which SMoS are suitable when analysing the safety of vulnerable road users (pedestrians and cyclists). The thesis attempts to answer this question using two different approaches: 1) a literature review focusing on existing surrogate measures and how well they consider vulnerable road users from a theoretical perspective, and 2) four observational studies which focus on the validity of SMoS and their relation to exposure.

The literature review focuses on identifying existing SMoS, and on two main aspects when evaluating their suitability for analysing the safety of vulnerable road users. Firstly, if the indicators theoretically are able to measure both the risk of collision and the potential for injury should a collision occur, and secondly, to what extent vulnerable road users were included in previous validation studies. The findings from the literature review are that the most commonly used indicators (Time to Collision Minimum and Post Encroachment Time) are also the most validated, but that they have several theoretical limitations, mainly that they to do not measure injury potential and that they measure the severity of an event based on the outcome rather than the initial conditions or potential/observed evasive actions. There are also several indicators which theoretically are more suitable but instead lack validation studies.

The observational studies, which make up the second part of this thesis, consist of an attempt at a large-scale validation study, followed by several studies which focus on the shortcomings discovered in the first attempt. The large-scale study is based on three weeks of video recordings made at 26 signalized intersections in seven European countries. The analysis of these videos resulted in three major findings. Firstly, the lack of comparable crash records made any large-scale validation attempts impossible. Secondly, the lack of comparability between the critical events identified by human observers and those identified by computer calculations made it infeasible to perform a long-term analysis. Thirdly, there is a significant relationship between meetings and critical events identified using Time to Collision Minimum and Post Encroachment Time, which suggests that some of the benefit of using those (and other indicators) might originate from their inherent connection to simple meetings between road users (i.e. exposure).

Following these results, the thesis presents a limited validation study based solely on the Scandinavian intersections followed by a suggestion for how a relative approach to validity might offer a potentially easier way of evaluating SMoS in the future. The results from these studies indicate that Time to Collision Minimum can measure safety to at least some extent, while Post Encroachment Time meassures it to a lesser extent. Due to the strong connection between critical events and meetings, the thesis also explores how a meeting between road users can be defined and how understanding what constitutes an opportunity for a crash might help to explain the so-called safety-in-numbers effect, as well as how future SMoS studies should consider meetings.

Key words: Surrogate Measures of Safety, Vulnerable Road Users, Traffic Safety, Video Analysis

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language**
ISSN 1653-1930 Bulletin – Lund University, Faculty of Engineering, Department of Technology and Society, 320		**ISBN** 978-91-7895-648-7 (print) 978-91-7895-649-4 (pdf)
Recipient's notes	**Number of pages:** 101	**Price**
	Security classification	

Signature

Date 2020-09-21

Surrogate Measures of Safety with a Focus on Vulnerable Road Users

An exploration of theory, practice, exposure, and validity

Carl Johnsson

Cover Photo by Hampus Norén

Faculty of Engineering
Department of Technology and Society

ISBN 978-91-7895-648-7 (print), 978-91-7895-649-4 (pdf)
ISSN 1653-1930

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Printed matter 3041 0903

Table of Contents

1. Populärvetenskaplig sammanfattning

Denna avhandling handlar om att undersöka hur nästan-olyckor kan användas för att studera trafiksäkerhet med ett fokus på oskyddade trafikanter (cyklister och gående). Den grundläggande idén är att farliga situationer i trafiken kan användas för att studera säkerhet, vilket kan göra det möjligt att snabbt och effektivt analysera säkerheten utan att först behöva vänta på olyckor. Avhandlingen fokuserar på att undersöka hur dessa farliga situationer kan identifieras baserat på videoinspelningar av 26 signalerade korsningar i 7 europeiska länder.

Idén om att studera nästan-olyckor har använts i olika former sedan 1960-talet och det finns därför en mängd olika metoder för hur man kan identifiera farliga situationer, samt flera olika förklarningar om vad som innefattar en farlig situation. Ett av resultaten från denna avhandling visar att *tid till kollision* är det vanligaste sättet att bedöma hur farligt ett möte mellan två trafikanter är. *Tid till kollision* är en indikator som uppskattar hur lång tid det kommer att ta tills att två trafikanter kolliderar förutsatt att båda trafikanterna fortsätter sin färd utan att bromsa. Denna avhandling fokuserar på hur bra mått som detta fungerar och om de verkligen kan användas för att studera trafiksäkerhet.

Resultatet från avhandlingen har ett tydligt budskap: måttet *tid till kollision* kan till viss del användas för att analysera säkerhet och producerar bättre resultat än andra indikatorer som testats i avhandlingen. Dock tyder resultatet även på att det finns en stark koppling mellan de testade måtten och antalet möten som sker. Det är till viss del självklart att det finns en sådan koppling, då det är omöjligt att ha en nästan-olycka mellan två trafikanter utan att också ha ett möte mellan dem. Samtidigt är det viktigt att nästan-olyckorna ger oss mer information om säkerheten än bara antalet möten. Tanken bakom nästan-olyckorna är att de inte bara ska mäta hur många möjligheter till olyckor som sker (möten) men också kunna identifiera de mest allvarliga mötena.

Resultatet från avhandlingen tyder alltså på att de indikatorer som testats har en stark koppling till antalet möten som sker. Den mest sannolika förklaringen är att indikatorerna misslyckas med att enbart identifiera farliga situationer, och även identifierar en mängd situationer som datorn felaktigt har bedömt som farliga. Detta kan antingen ske på grund av hur indikatorerna är designade, eller hur de har beräknats i datorn.

Avhandlingens resultat är intressant i samband med den snabba utvecklingen och framväxten av videoanalys som skett under 2010-talet. Videoanalys och datorseende har sett en markant förbättring med hjälp av metoder som maskininlärning. Detta skapar nya möjligheter för att studera trafikbeteende med hjälp av modern teknik. Resultatet och medföljande diskussioner om möten och nästan-olyckor i denna avhandling har möjligheten att förbättra hur den nya tekniken kan användas för att bättre studera trafiksäkert med fokus på oskyddade trafikanter.

2. Introduction

Surrogate Measures of Safety (SMoS) are meant to function as tools to investigate traffic safety. The term *surrogate* indicates that these measures do not rely on crash data but instead are meant to be an alternative and a complement to analyses based on crash records. The *traffic safety* can generally be considered as *the absence of unintended harm to living creatures or inanimate objects* (Evans, 2004).

The underlying idea of SMoS is that there are some critical events in traffic that, while not resulting in a crash, are somehow more severe than normal events. The idea is that these *critical events* can be used to study traffic safety without directly relying on the crashes themselves. SMoS generally works by categorising traffic events based on some measure of severity. Assuming more severe events have a stronger causal connection to crashes and therefore safety, an analysis of the safety can be made by investigating the occurrences of severe events. It is also possible to study the causal mechanisms that result in severe events, which in turn can allow for a better understanding of why crashes occur.

The approach based on SMoS has several advantages compared to accident-based analysis. The main advantage is that the analysis is pro-active (there is no need to wait for crashes), and in some conditions more time-efficient, informative, and even accurate (Hydén, 1987; Å. Svensson, 1992). As a further consideration, SMoS can also be applied in cases where crash records are lacking, or in other scenarios in which more traditional methods are of limited use.

Background

Need for SMoS and some limitations of crash data

Traditionally, road safety is described in terms of number of crashes or injuries that occur in traffic. While crash data have the most direct connection with traffic safety, it has several limitations:

- Crashes are rare and random events, and the number of crashes recorded every year at the same place is not the same, even if the traffic situation is unchanged. Following this perspective, the number of crashes per year is also a somewhat indirect measure. The *true* safety characteristic of interest is the *expected number of crashes per year* that cannot be directly measured but must be estimated based on the crash history or using other methods (Hauer, 1997a).
- Crashes are rare, and it takes a long time to collect enough crash data to produce reliable estimates of the expected number of crashes. During that period the traffic conditions might (and usually do) change. There is also an ethical problem in that one must wait for enough crashes to occur and thus for people to suffer before anything can be said about the (un)safety.
- Not all crashes are reported. The level of underreporting depends on the accident's severity and types of road users involved. This is especially a problem for Vulnerable Road Users (VRUs) (Alsop & Langley, 2001; Amoros et al., 2006; R. Elvik et al., 2009).
- The rarity of crashes makes them difficult to directly observe. Accident reconstructions and in-depth investigations are usually costly and not always possible to perform. It is therefore difficult to gain a good understanding of the process preceding an accident using solely crash data.

Using SMoS offers an alternate method which can allow for traffic safety analyses in scenarios in which crash data is either completely lacking or cannot properly provide a sufficient safety analysis.

SMoS from a theoretical perspective

The basic concept of SMoS theory is that traffic consists of a number of elementary events. These events differ in their degree of severity (unsafety), and there exists some relation between the severity and frequency of events of that severity. Hydén (1987) illustrated the concept with a *safety pyramid* (see Figure 1).

The base of the pyramid represents the frequently occurring and safe "normal" passages. The top of the pyramid consists of the most severe events such as crashes resulting in fatalities or injuries. If the shape of the relation between the severity and frequency of the events is known, it is theoretically possible to estimate the frequency of the very severe but infrequent events (crashes) based on known frequency of the less severe, but more easily observable events (critical events).

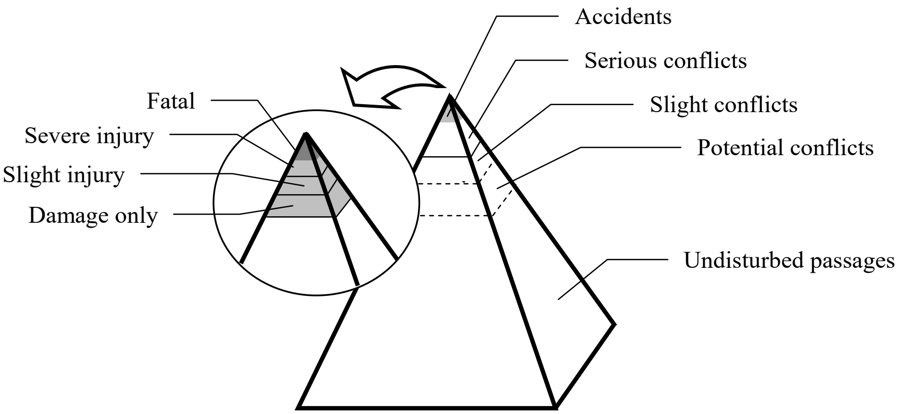
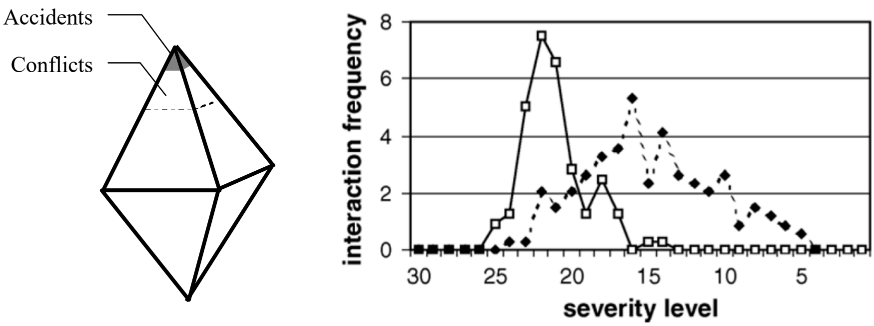


Figure 1. *Safety pyramid* (Hydén, 1987).

Å. Svensson (1998) describes a slightly altered “pyramid”, in which events without any risk of a collision (such as single passages) are excluded. Elaborating further on the meaning of the severity distribution shape, she pointed out that the most frequent events are not necessarily the safest ones (*severity diamond* model, see Figure 2a).



a) *b)*

Figure 2. *Severity diamonds* (Å. Svensson, 1998): a) conceptual representation; b) observed distributions of events with different severity levels (according to the Swedish traffic conflict technique) at two sites.

Moreover, comparing different types of road environment, Å. Svensson (1998) showed that the shape of the distribution varies depending on factors such as regulation form, road design, frequency of interactions, type of manoeuvre, and road users involved, etc. (Figure 2b).

What is severity?

The concept of *severity* also requires clarification. Most SMoS express the severity of an event as its proximity to a collision in terms of time or space (Zheng et al., 2014c). However, the proximity to a collision is only one dimension of *severity*. Another dimension of *severity* is the potential consequences had a collision occurred (Laureshyn, 2010). The framework described by A. Laureshyn et al. (2010) provides a more complete overview of *severity* by dividing the concept into two categories – collision risk and injury risk of an event – and arguing that severity could be estimated by combining these two aspects. This division makes it possible to differentiate between the factors affecting collision risk and those influencing injury risk as shown in Table 1.

Table 1. Factors affecting collision risk and injury risk respectively. Based on Svensson (1998).

Collision risk	**Injury risk**
• Closeness in time • Closeness in space • Speeds of the involved road users	• Speed differences • Mass differences • Relative angle • Fragility of the involved road users

Relation between critical events and crashes

How the events of different severity are related has a direct effect on whether there are theoretical grounds to extrapolate the knowledge from the less severe events to the more severe ones and finally, crashes.

Two alternative models relating critical events and crashes have been described by Güttinger (1982). In the first, the critical events are defined as a set of initial conditions that, depending on the successfulness of the evasive action, either develop further into a collision, or resolve without any consequences (see Figure 3*a*). Defined this way, critical events (called conflicts in the figure) and collisions belong to the same dimension, as the first always precedes the second, and a critical event can, with a certain probability, develop into a collision. In the alternative model (Figure 3*b*), it is the evasive action that results either in a collision or *an avoided collision* – a critical event. In this definition, critical events and collisions exist *in parallel*.

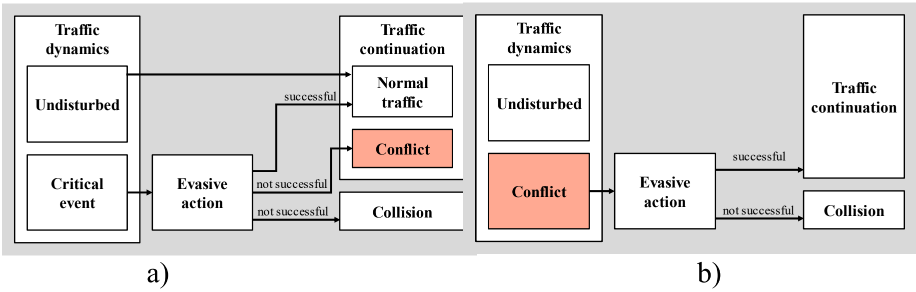


Figure 3. Two models of relation between critical events and crashes (adapted from Güttinger (1982)): a) conflict precedes a collision; b) conflict is *parallel* with a collision.

Which model lies behind a SMoS is important. If critical events and crashes do not belong in the same continuum, the use of critical events to predict frequency of crashes is not well-motivated. For example, there might be some factors always present in collision situations and absent in critical situations (or vice versa) that are crucial for whether the situation is resolved successfully or not.

Davis et al. (2011) suggest an alternative model to understand the causal mechanics between traffic events and crashes. Their model (Figure 4) outlines that the probability for a traffic event to develop into a crash depends on two conditions. In this model, traffic events can be explained by a set of initial conditions [U] and a set of possible evasive actions [X]. The outcome [Y] is dependent on both the initial conditions and the possible evasive actions. Indicators that measure the initial conditions identify critical events based on the closeness of the involved road users, using metrics such as the physical distance between road users or the time separating two road users. Indicators that measure evasive actions identify critical events based on the magnitude of any evasive action, using metrics such as braking, running, or swerving (Davis et al., 2011). A SMoS should ideally reflect both aspects of the model to accurately estimate a collision risk.

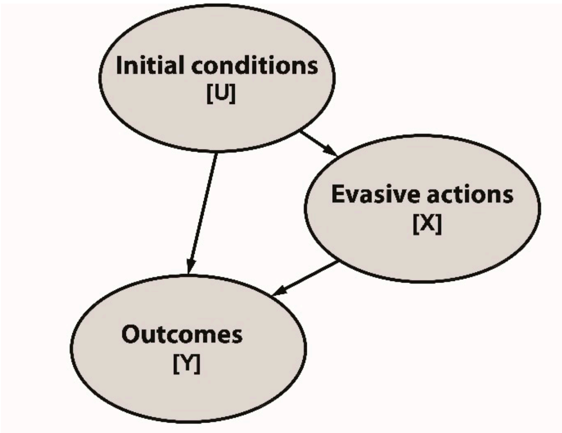


Figure 4. Causal model. Adapted from Davis et al. (2011).

Validity

Validity is a crucial aspect of any study or method. Validity, in general, relates to the approximate truth of an inference (Shadish et al., 2002). Validity is not necessarily a matter of *yes or no*, but a matter of degree (Carmines & Zeller, 1979); whether a certain level of validity is considered *sufficient* is therefore usually rather a matter of argumentation, debate, and agreement, than a measurable aspect that should exceed a certain threshold. Validity has to be assessed relative to purposes and circumstances (Brinberg & McGrath, 1985).

The validity of SMoS concerns the crucial severity distribution described in the previous section. The main aim of developing and using SMoS is to measure traffic (un)safety; therefore, validity of an indicator means to what extent it describes (un)safety. There are potentially several ways to study the validity of a SMoS – expert judgements, comparison with other *indirect* measures, comparison with observed/reported/estimated crashes, etc. The *strength* of validation will vary depending on which approach is used.

Previous attempts at making this kind of validation study have used several different methods aimed at analysing the relation between critical events and crashes. The list below briefly describes several approaches:

- linear correlation between observed critical events and recorded crashes, e.g. Baker (1972),
- minimisation of the variance of the ratio between crashes and critical events, e.g. Hauer and Gårder (1986); Hydén (1977),
- linear correlation between critical events and the expected number of crashes calculated from a flow-based crash model (Lord, 1996),

- estimation of the expected number of crashes from a critical event-based crash model (El-Basyouny & Sayed, 2013),
- comparison of the expected number of crashes estimated from the crash history, with the expected number of crashes estimated from an extreme value theory approach using critical events (Songchitruksa & Tarko, 2006)
- comparison between a critical event-based and crash-based before-and-after study (Sacchi et al., 2013)

Reliability of measurements

The concept of reliability refers to the accuracy and the consistency of measurements – in other words, a measured value should very closely represent the *true* value, and the measurement error should remain unaffected regardless of measurement location, time of day, traffic situation, etc., thus ensuring that measured differences reflect the actual difference in the studied phenomenon and not in the measurement's accuracy (Laureshyn, 2010).

There are several aspects to reliability from the perspective of SMoS. For example, the accuracy of measurements for individual traffic events (road users' position, speed, etc.) and the detection errors related to that. There is also the question of the observation period necessary to collect enough events to be able to generalise their frequencies (e.g. estimate the *expected number of critical events*).

The first point can be further divided into two categories: human observers and automated data collection. Human cognitive capacity puts significant limitations on the complexity of the analysis that is feasible to perform in field conditions and in real time. Consequently, the techniques use very few severity categories and are often based on qualitative rather than quantitative classifiers. When it comes to human estimates of quantitative measures, several studies show that with proper training, it is possible to get adequate accuracy (Hydén, 1987; van der Horst, 1984).

The automated data collection methods are objective per definition, but the technical details and the performance of the system might influence its result. In cases of automated video analysis, such factors, beside the choice of the video processing algorithm itself, are (Morse et al., 2016):

- Quality of the underlying calibration;
- Characteristics of the camera (e.g. resolution) and characteristics of the installation (height and angle);
- The complexity of the traffic scene;
- Environmental conditions (e.g. weather and darkness).

As for the question of the necessary observation period, there is very limited research on how long a SMoS study should be. A study conducted by Hauer (1978) offers some insight into how accuracy of the estimated *expected rate of critical events* improves with the extension of the observation period, which can be used for decisions on how long of an observation period *is long enough*. However, the frequency of critical events is highly dependent on which indicator is used and at what threshold of severity an event becomes “critical”. Therefore, there are no general guidelines for how long a SMoS study should be conducted.

Exposure

Critical events are not merely a measure of exposure (Hauer, 1982). The purpose of exposure is to *take account of the amount of opportunity for crashes* (Chapman, 1973), while *critical events*, similar to the actual crashes, are a result of both the exposure and the crash risk at a given site. It is common to define exposure in terms of number-of-vehicles per time-unit, or vehicle-kilometres travelled. However, it is also possible to use a definition in which a unit of exposure is actually an event that can be seen as an opportunity for a crash to occur (Elvik, 2015).

An event-based measurement of exposure has some advantages compared to typical traffic counts when used in conjunction with SMoS. Both crashes and *critical events* are subsets of a larger set containing all events with any probability of a crash, i.e. the event-based exposure. Figure 5 illustrates this relationship between elementary units of exposure, crashes, and *critical events* at different threshold levels.

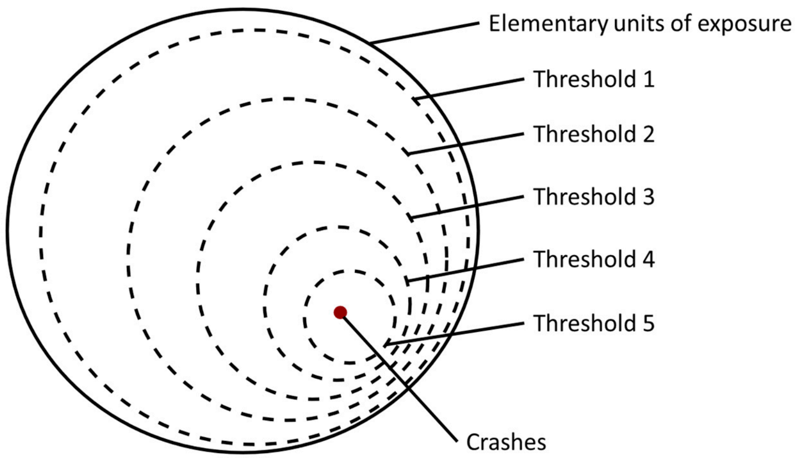


Figure 5. The theoretical relation between elementary units of exposure, SMOS at different threshold values, and crashes. Adapted from Amundsen and Hyden (1977, p. 137).

The event-based definition of exposure provides a clear and logical link between the exposure, risk (probability of an exposure unit developing into a crash), and crashes. It also allows controlling for the exposure in SMoS studies in a transparent way. To explain this, imagine a validation study in which a certain definition of a *critical event* shows a strong correlation to crashes. The question should be whether this correlation is a result of causal relations between the *critical events* and crashes or is simply a result of the fact that both are a subset of elementary events constituting the exposure. As illustrated in Figure 5, if the threshold used is lenient, it would include a great number of elementary traffic events and as such, will not be much different from the exposure. Still, it might appear as a well-functioning definition of a *critical event* just because there is a strong relation between crashes and exposure.

Research Gaps

There are several unresolved issues when it comes to SMoS, such as: selection of the appropriate indicators, the validity of the indicators, data collection, and analysis procedures, etc. The traditional reliance on trained human observers is also a potential obstacle for the method. Recent computer science research has been characterized by great improvements in sensor technologies which can be applied for collection of traffic data in general and SMoS in particular (Laureshyn, 2010; Saunier et al., 2010; Tarko et al., 2017).

However, it is unclear to what extent the previous research based on human observations can be applied to analyses made using Computer Vision.

Researchers new to the SMoS struggle to gain a clear overview of the current state of the field. The literature in this domain is expansive and stretches back to the 1960s, and the technical improvements in the field have recently seen a rapid growth. The lack of a comprehensive overview of the field can potentially result in old research being repeated or the wrong type of surrogate measure being applied. This lack is especially relevant for studies of Vulnerable Road Users (VRUs); most of the earlier research has been performed with a distinct focus on cars, and the existing methods might not be directly applicable on pedestrians or cyclist

Aim

The main research objective of this thesis is to investigate **how Surrogate Measures of Safety should be used to study Vulnerable Road Users**. This objective has been divided into the following research questions, which will be answered throughout the thesis.

1. Which indicators exist and how suitable are they for VRU studies from a theoretical standpoint?

2. To what extent have the various indicators been validated, and to what extent have these efforts included VRUs?

3. How are SMoS studies generally conducted, in terms of observation length, complementary factors such as other behavioural observations, and exposure etc?

4. How well do the various indicators reflect safety, and which threshold values produce the best results?

5. How should the various indicators be validated?

6. How should exposure be considered in SMoS studies?

Structure of the thesis

The structure of the thesis is divided into two main parts. The first part presents a comprehensive literature review of SMoS, while the second part focuses on several observational studies. Both parts aim to provide an answer to the main research objective. However, the first part focuses on the first three questions, and the second part focuses on the last three. The second part is further divided into a primary study which results then function as the basis for the remaining studies.

This thesis incorporates the following 5 papers:

- Paper 1 - In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators.

- Paper 2 - Validation of surrogate measures of safety with a focus on cyclist-motor vehicle interactions

- Paper 3 - A relative approach in validation of surrogate measures of safety

- Paper 4 - The *safety in density* effect for cyclists and motor 1 vehicles in Scandinavia: An observational study

- Paper 5 - A general method for identifying evasive actions and calculating time to collision from trajectory data

These papers relate to different parts of the thesis: Paper 1 is mostly covered in the first literature part of this thesis, Paper 2 is the primary study of the second part of the thesis, and the following Paper 3, 4 and 5 rely on the results and conclusions of Paper 2.

The papers and the thesis itself are based on the work and outcome of the EU project In-Depth Understanding of Accident Causation for Vulnerable Road Users (InDeV) (HORIZON 2020, Project No. 635895). Consequently, some of the papers share data and general methodology.

PART 1

3. A literature review of SMoS

Surrogate Measures of Safety were applied for the first time half a century ago (Perkins & Harris, 1967), and their underlying theories and applications have evolved over the years. The last decade especially has been characterized by great improvements in sensor techniques and computer vision, which can be applied for the collection of traffic data in general and surrogate safety measures in particular (Laureshyn, 2010; Saunier, Sayed, & Ismail, 2010; A. P. Tarko, Romero, Bandura, Ariyur, & Lizarazo, 2017).

The available literature in the field of Surrogate Measures of Safety is vast and diverse, and a strong increase in interest for the technique in recent years can be observed. The vastness and diversity of the literature in the field make it difficult to discern dominant practices and customs in research that applies surrogate measures of safety. This makes it challenging for researchers new to the field to gain a clear view of the scientific state-of-practice, and even for more experienced researchers there is a risk of losing track of critical points of attention. The lack of overview seems to lead to *reinventing the wheel* and errors from the past being repeated. Therefore, a comprehensive scoping literature review has been conducted, analysing the literature in the field in a systematic and structured way. The aim of this review is to analyse the literature from two distinct perspectives:

- The safety of Vulnerable Road Users (VRU) and how different SMoS indicators are suited for such a task, and how many validation studies included VRUs.

- A more general and holistic focus on which indicators have been used to analyse road user interactions and how they have been used to study traffic safety in general.

Method

A systematic and transparent protocol was set up to find relevant studies. The main method for locating the literature for this review was searching the following databases: ScienceDirect, TRID, Web of Science, Engineering Village and Scopus. Applied search terms were: *traffic conflict, traffic conflict techniques, surrogate safety, safety critical event, indirect safety, near-accident* and *near-miss*, combined with the terms *traffic* and *traffic safety*. For practical reasons, the studies had to be written in English, Swedish or Dutch. Snowballing was also used for references deemed of high importance based on the reference list of found literature and on the authors' previous knowledge about surrogate measures of safety. Older studies were included without age restrictions as long as they could be retrieved. All publications up until the end of 2015 are included.

The original search yielded 2445 hits. Thorough screening was performed; only studies that made use of site-based observations (i.e., data collected from real traffic events, using one or more cameras or sensors that remain at the same position for a certain period of time) were included. This excluded, among others, SMoS data collected from driving simulators, microsimulation models, and naturalistic driving. After removal of duplicates, studies that were outside the scope, and publications for which no full text could be retrieved, a total of 239 publications could be included in this scoping review.

Description of SMoS

The following overview summarises the different surrogate indicators of safety that were found in the literature search. Due to numerous variations of similar indicators, some indicators are combined into categories, containing the main indicator and various alternatives based on the same concept. Any previous validation studies found in the review are also presented with each indicator.

Time to Collision

Time to Collision (TTC) is a measurement that calculates the time remaining before the collision if the involved road users continue along their respective paths. This continuous variable can only be calculated while the road users remain on collision course. The two most commonly used indicators based on TTC are TTC_{min}, which is the minimum TTC value calculated in an event, and Time to Accident (TA), which is the TTC value at the beginning of an evasive action. Usually, both indicators use a threshold value to differentiate between severe and non-severe events.

Other indicators based on, or similar to, TTC are: Time to Zebra (TTZ) (Várhelyi, 1998), the Time-to-lane Crossing (TTL), and the reciprocal of TTC (i.e. 1/TTC) (Chin et al., 1992). There are also some indicators that use TTC values from several moments in time. Minderhoud and Bovy (2001) propose Time-exposed TTC (TET), which is the time during a meeting when the TTC is below a certain threshold value, and Time-integrated TTC (TIT), which is the area between the threshold level and the TTC curve when it drops below the threshold.

Lastly, the T2 indicator, suggested by A. Laureshyn et al. (2010), is the predicted arrival time of the second road user, calculated while the first road user has not yet left the conflict point. When the road users are on a collision course, T2 is equal to TTC.

Five studies attempted to validate TTC-based indicators. While all the studies use different methods to evaluate the relationship between critical events and crashes, they all find a strong correlation between the two. Four out of five studies include VRUs to some extent; however, only the study by Lord (1996) explicitly focused on pedestrians, while the report by Hydén (1977) separates the result for VRUs and motor vehicles. The studies by El-Basyouny and Sayed (2013) and Sacchi and Sayed (2016) both include VRUs, but only 4.6% of all critical events they use include VRUs. Note that both of these studies use older data described by Sayed and Zein (1999).

Post-encroachment Time

Initially introduced by Allen et al. (1978), the Post-encroachment Time (PET) is calculated as the time difference between the moment the first road user leaves the path of the second, and the moment the second reaches the path of the first (i.e. the PET indicates the extent by which they missed each other).

Indicators similar to PET include Gap Time (GT), which is the time between the entries into the conflict spot of two vehicles, and Encroachment Time (ET), the time when the first vehicle entering the conflict area infringes on the predicted path of the second vehicle (Allen et al., 1978).

Proposed by Hansson (1975), the Time Advantage (TAdv) can be considered an extension of the PET concept (A. Laureshyn et al., 2010). The TAdv is the predicted PET value, provided that the road users continue with their paths and speeds. Alhajyaseen (2015) suggests the Conflict Index (CI) indicator, which combines PET with the speed, mass, and angle of the involved road users to estimate the released kinetic energy in a collision.

In total, seven studies with eight attempts to investigate the PET's validity were identified. Similarly to TTC, most of the studies indicate a correlation between critical events and crashes, with the notable exception in the study by Lord (1996), in which the PET definition showed no correlation and was discarded from further study. Furthermore, the studies that used Extreme Value Theory (Songchitruksa & Tarko, 2006; Zheng et al., 2014a, 2014b) all focused primarily on the method and all noted that further research is required to achieve more reliable results.

Deceleration

There are several deceleration-based indicators that describe the severity of a traffic situation. The Deceleration Rate (DR) or the initial DR quantifies the magnitude of the deceleration action of a driver the moment he or she begins an evasive braking manoeuvre. Additionally, the Deceleration to Safety Time (DST) describes the nearness to a collision by calculating the necessary deceleration for a driver to stop being on a collision course (Hupfer, 1997).

Tageldin et al. (2015) suggest that the Jerk Profile (the time derivative of acceleration) and the Yaw Rate (the angular velocity of the road users' rotation) can be used for identifying evasive actions by motorcyclists. The Jerk Profile estimates the intensity of the braking action by observing the change in acceleration. The Yaw Rate quantifies the swerving behaviour of the motorcyclists by calculating the change in the heading angle of the motorcycle. Combining these two indicators allows for an estimation of severity for both braking and swerving manoeuvres.

No previous validation studies were found for deceleration-based indicators.

Other indicators

Several indicators do not fit into any of the indicator categories presented above. These indicators estimate severity differently, but they have all been suggested as alternatives to the more commonly used indicators. No previous validation studies were found for any of these indicators.

Tageldin and Sayed (2016) suggest that evasive action-based indicators, such as pedestrians' step frequency and step length could be used to identify severe events involving pedestrians. Cafiso et al. (2011) describe the Pedestrian Risk Index (PRI), which combines the TTZ with driver reaction times and the braking capabilities of the vehicle to estimate the risk of collision and its potential severity.

Bagdadi (2013) defines Conflict Severity (CS) as a combination of the indicators DeltaV, TA, and an assumed maximum average deceleration of a vehicle. The DeltaV indicator measures the change in velocity forced on the road users by a collision depending on the speed, the mass, and the angle at which the road users approach each other (Shelby, 2011). Another indicator that also uses DeltaV is the Extended DeltaV indicator suggested by Laureshyn et al. (2017). It combines DeltaV with the T2 indicator and a deceleration constant to capture the nearness to a collision, as well as the potential consequences of an event.

Kuang et al. (2015) developed an indicator called the Aggregated Crash Index (ACI) based on the causal model presented by Davis et al. (2011). This indicator is meant for car-following scenarios and consists of four conditions in a tree structure. The conditions estimate both the initial conditions of an event and the potential for evasive action. The ACI is then calculated as the accumulation of the collision probabilities of all possible outcomes.

Ogawa (2007) discusses a space occupancy index based on personal space, which expresses the spatial sizes necessary to maintain road safety for pedestrians, bicyclists, and motor vehicles (MVs) An area around the road user is defined based on the characteristics of each road user type. The number of critical events is then estimated by the number of incursions into the road user's personal space that occur.

Traffic conflict techniques

To capture the severity of an event, indicators can be combined to provide a better understanding of the situation. The rationale behind this approach is that many indicators are not sufficiently universal and cannot be applied to all traffic events. It is plausible that various indicators represent partial images of the true severity of a traffic event (Ismail et al., 2011).

The most common examples of this approach are the traffic conflict techniques (TCTs) that were mostly developed in the 1970s and 1980s. A TCT is a framework for observation-based safety studies, including observation methods, instructions on how to use the technique, as well as a set of indicators used to identify conflicts (severe events). However, there are also some examples of indicator combinations outside of the well-defined TCTs. For example, Lu et al. (2012) combined the incomplete braking time and the TTC to calculate conflict severity. Wang and Stamatiadis (2014) used the required braking rate, the maximum available braking rate, and the TTC to create an aggregate crash propensity metric, and Ismail et al. (2010) integrated the DST, the PET and the TTCmin.

The following sections briefly present several TCTs.

The American conflict technique

The American TCT defines critical events as the occurrences of evasive actions, recognisable by braking and/or weaving manoeuvres (Parker & Zegeer, 1989). Five studies that focus on the validity of the American TCT were found. It is noteworthy that the study by Migletz et al. (1985) used an alternative version of the American TCT, which also included TTC_{min} to identify critical events. All studies found a strong correlation between critical events and crashes.

The Canadian conflict technique

The Canadian TCT uses TTC_{min} in conjunction with a subjective component named the Risk of Collision (ROC) to determine the severity of an event. Three levels of severity are used for both TTC_{min} and ROC, and the final severity is estimated by adding them together. The ROC levels are low, moderate, and high, and the TTC_{min} levels are less than 2 s, less than 1.6 s, and less than 1 s, respectively (Brown et al., 1984).

Three studies focus on the validity of the Canadian TCT. Note that the Brown (1994) study uses the TA indicator instead of the TTCmin. The validation studies all found a clear correlation between critical events and crashes. However, while all the studies included VRUs to some extent, none of them focused on VRUs.

The Dutch conflict technique

The Dutch Objective TCT for Operation and Research (DOCTOR) defines a critical event as a situation in which a collision is imminent, and a realistic probability of personal injury or material damage is present if no evasive action is taken.

The severity of a conflict is estimated in several steps. First, a general subjective severity is made. The event is then broken down into the probability of collision and the potential injury severity. The probability of collision is calculated by using the TTC_{min} value or the PET value, and the potential injury severity is estimated subjectively (Kraay et al., 2013).

Two studies focused on the validity of the Dutch TCT were found. Both studies indicate a significant similarity between critical events and crashes. While the study by van der Horst et al. (2007) only focused on motor vehicles, the study by van der Horst et al. (2016) did have a focus on pedestrians. It is also noteworthy that the study by van der Horst et al. (2007) is one of the only two studies found in this review that attempted some form of process validation. The study compared observed critical events from four different locations with video-recorded crashes from the same locations.

The Swedish conflict technique

The severity of a traffic event in the Swedish TCT is calculated based on a combination of TA and conflicting speed (the speed of the road user at the moment an evasive action starts). This means that a lower TA value is considered more severe if the involved road users enter the situation with a higher speed (Hydén, 1987).

Three studies investigate the validity of the Swedish TCT, including the second attempt of a process validation by Hydén (1987), in which the TA values and the conflicting speed of critical events are compared to the same values gathered from in-depth studies of crashes. All three studies indicate a significant relationship between critical events and crashes for VRUs. The studies by Hydén (1987) and Svensson (1992) included a separate analysis for VRUs, and the work by Shbeeb (2000) focused specifically on events involving pedestrians.

TCTs using only a subjective severity rating

The Canadian TCT and the Dutch TCT both use a combination of an objective indicator and a subjective severity rating. However, some techniques rely solely on a subjective severity rating to identify critical events. These TCTs are the British (Baguley, 1984; Kaparias et al., 2010), the French (Muhlrad & Dupre, 1984), the German (Erke, 1984)), the Austrian (Risser & Schutzenhofer, 1984), and the Czech techniques (Kocárková, 2012). These techniques use several predefined, subjective severity grades to identify critical events, that are often based on the closeness of road users and the occurrence of uncontrolled evasive actions.

Three studies investigate the validity of the British TCT. All three studies found strong correlation between critical events and crashes, but none of them included conflicts involving VRUs.

An overview of the usage of SMoS

This section will present how SMoS have been used in practise, including the usage of different indicators and other factors, such as the number of observed locations and the duration of observation. Figure 6 shows the distribution of the found publications over time. The graph shows both the publications that are included in the study and those identified as potentially relevant (based on the abstract of the paper), but for which no full text could be retrieved. The earliest application of SMoS included in the review was done by Perkins & Harris (1967). From that starting point the usage of SMoS became more popular throughout the 1970s and 1980s. After the early parts of the 1990s, the usage then diminished somewhat, which continued until the steep increase which can be seen starting around 2010.

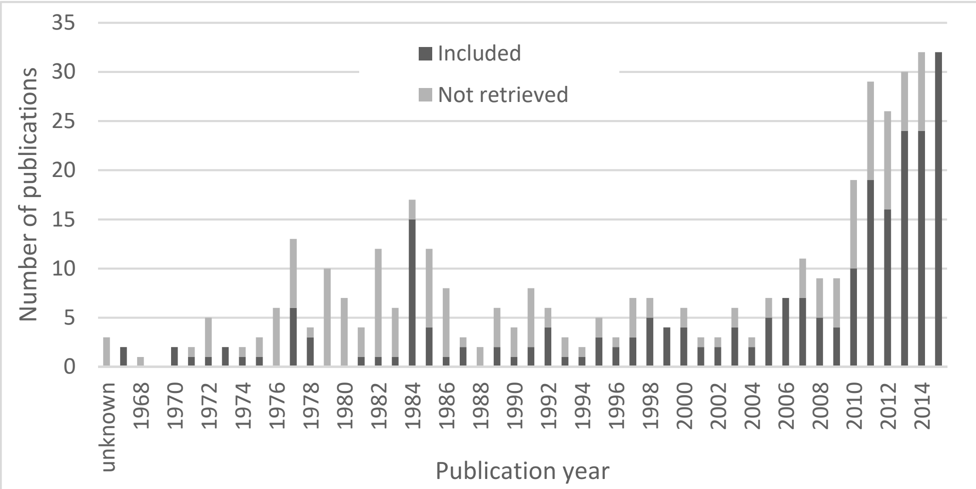


Figure 6 – Distribution of publications over time.

A potential explanation for the downturn during the 90s is that up until that point, mostly TCTs using human observes were used in practise (Asmussen, 1984). It is possible that this became increasingly expensive compared to other methods during that period. The steep increase around 2010 corresponds with both recent improvements in advanced video analysis techniques (Laureshyn, 2010; Saunier et al., 2010) and an increased focus on VRUs, which are generally harder to study using crash data.

Indicators and traffic conflict techniques

Figure 7 shows the frequency of which indicators are used in the publications and Figure 8 shows the frequency of which TCTs are used. Note that a separation has been made between studies before and after the year 2005 to gain a better understanding of the current state of use.

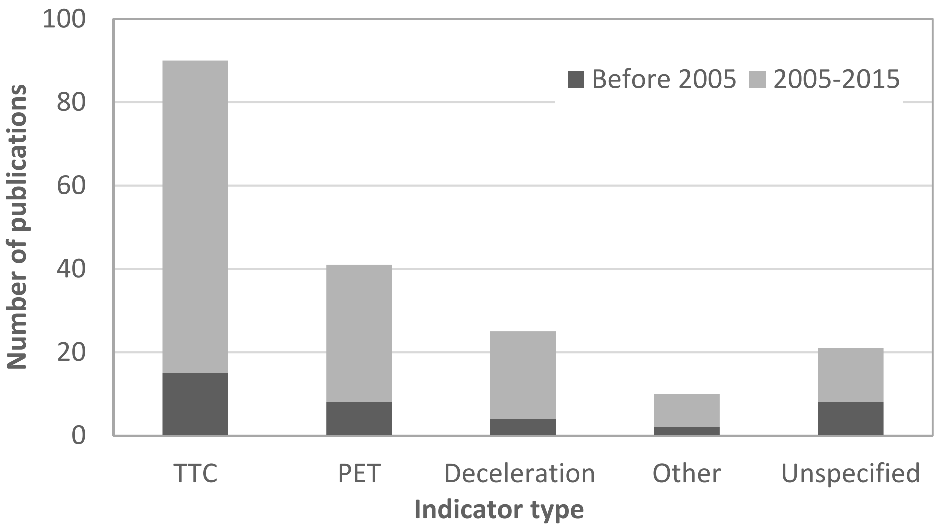


Figure 7 – Usage of indicators.

The main result from Figure 7 shows that indicators based on TTC are the most common by a significant margin (most of these studies use TTC_{min}, and only a few use other variants such as TA). PET is the second most common indicator, followed by the various indicators based on deceleration. The category *unspecified* relates to papers that provide insufficient details to identify the indicator(s) or technique(s) that have been applied. Most of these papers made use of some abstract, unspecified concept of *critical events* without explaining what parameters or thresholds were used to select them. Figure 8 shows that the US TCT and the Swedish TCT have been most commonly applied. When only looking at the last decade, however, the US TCT has been used less frequently than the Swedish TCT. Comparing the two graphs shows a clear result – since 2005, a shift away from the usage of TCTs have occurred and other indicators, mainly TTC and PET, have become much more common.

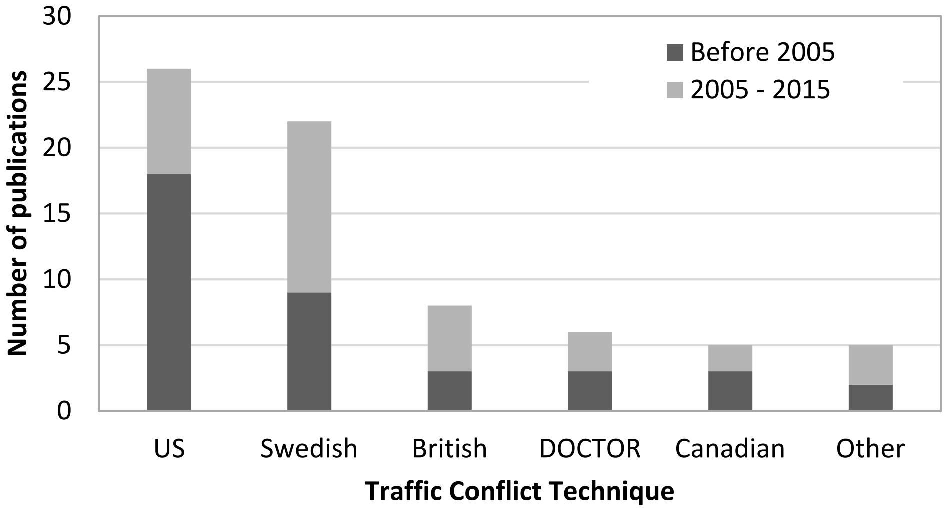


Figure 8 - Usage of traffic conflict techniques.

Figure 9 shows the number of observation sites in the studies and Figure 10 shows the average duration of the observations per site. Many studies (approximately one third) take place at only one observation site. On the other hand, another third of the studies make observations at 5 or more sites. Relatively short observation periods per observation site are also quite common. 45% of all studies observed less than 8h per site, while only 22% of all studies observed for more than 24h per observation site. Also note that 23% of the included publications did not include any information about the duration of the observations at all.

One might expect that there is a trade-off between the number of observation sites and the duration of observation at each site, however, such a trade-off is barely discerned. For instance, when looking at studies that observed only one site, short observation times of less than 8h and less than 4h are as common as in multi-site studies (45% and 28% of the single-site studies compared to 45% and 22% for all studies taken together, respectively). Only studies with a duration of more than 24h per site are somewhat more common in the subgroup of single-site studies (30% for single-site studies, compared to 22% for all studies taken together). There are also noteworthy differences before and after 2005. Since 2005, significantly more studies have focused on a single location. Likewise, there are more studies after 2005 with a shorter observation period.

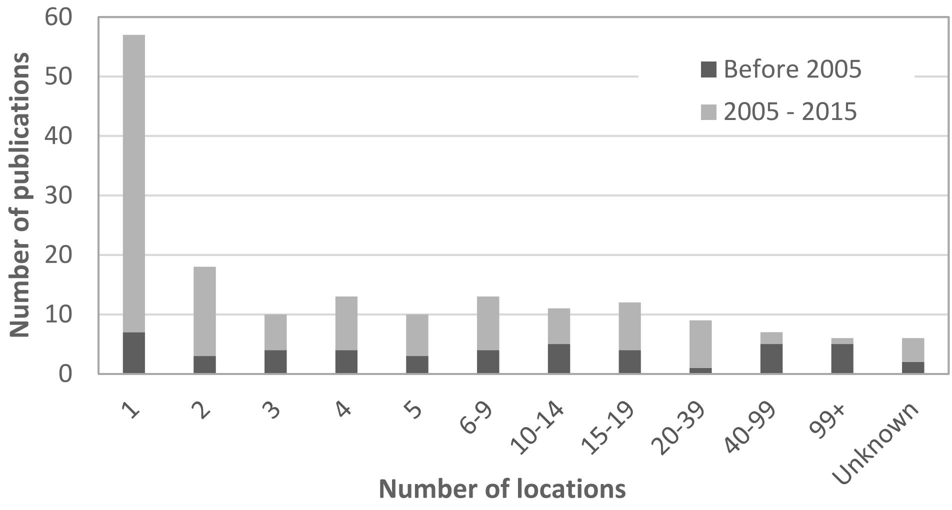


Figure 9 – Number of observed locations.

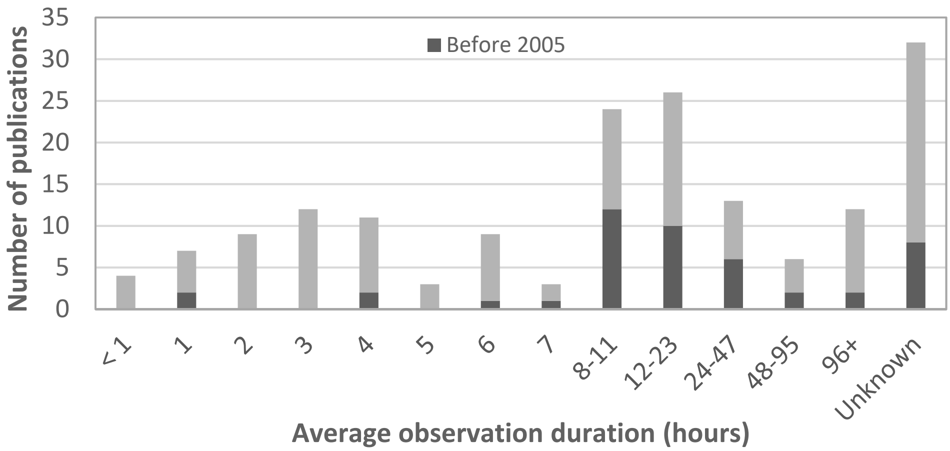


Figure 10 – Average duration of observation per site.

Data collection method

The methods used to collect SMoS are presented in Figure 11. Different forms of manual observation (the three left-most bars) have been most common over the years. However, the number of publications that apply video analysis software take up a significant share as well, especially after 2005. Fully manual observations (i.e. human observers on-site without video support) are a relatively large category for all publication years taken together but have rarely been applied in recent years. It is also noteworthy that manual observation from video (i.e. without a human observer on-site) is the largest individual category of data collection methods and is very common since 2005.

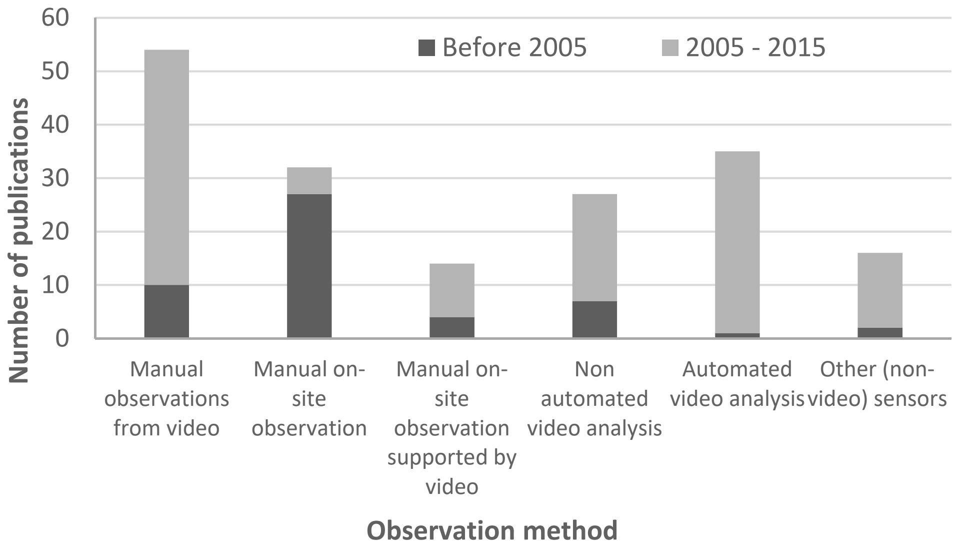


Figure 11 – Data collection methods.

Additional data

Any additional data that are collected in conjunction with SMoS data are shown in Figure 12. Only a few studies do not include any additional data together with SMoS data. This indicates that results from SMoS are not usually applied as a standalone safety analysis but are more used as a complement and in conjunction with other methods.

The most common additional data that is collected is exposure data (mostly measured using traffic counts taken during the SMoS data collection). Somewhat less commonly collected are crash data, information about the infrastructure, and systematic behavioural observations. The category *Other* is fairly large as well, including very diverse types of data such as results from microsimulation or driving simulator studies, road user characteristics such as gender and age, and survey or interview data.

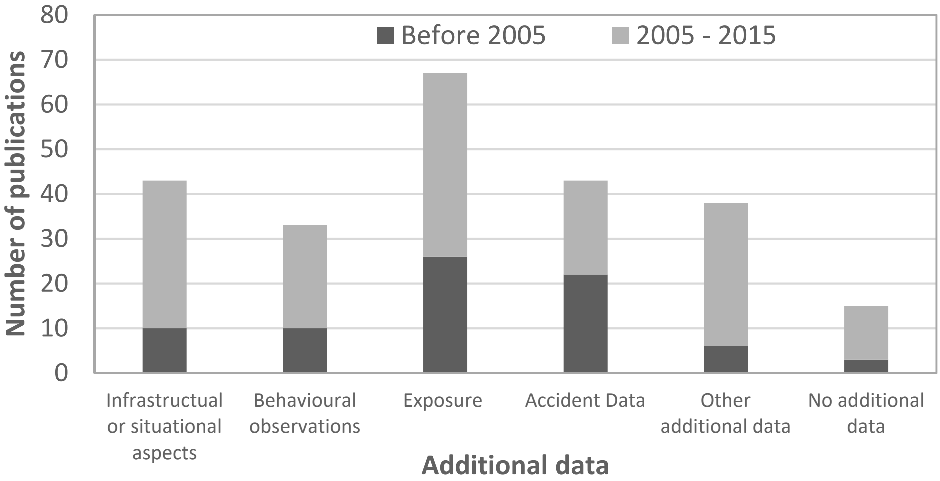


Figure 12 – Additional data collected in SMoS studies.

Suitability of SMoS for studying VRU's safety

SMoS can be used in various settings involving different road users. However, not all indicators are necessarily equally suitable for all kinds of settings and events. Three main aspects of a surrogate safety indicator can be used to discuss the suitability of different indicators in various settings. First, an indicator should include the theoretical aspects important for different settings. Second, it should have robust validity, and thirdly, reliability. Based on the theoretical perspective presented in the introduction of this thesis (page 10), the severity of an event can be estimated with a combination of injury risk and collision risk. A collision risk can be further reduced into a combination of initial conditions and evasive actions, meaning that the collision risk can be estimated using either or both aspects.

An indicator should preferably reflect both parts of collision risk, as well as injury risk, in all settings where the indicators are used. For example, an indicator that is used to study VRUs' safety should preferably be able to consider the nearness of the road users, the potential evasive actions of the involved road users, and the fragile nature of the VRUs.

Table 2 provides a summary of all indicators described in this chapter from this theoretical perspective. It shows whether or not the indicators reflect each of the ideal requirements for a surrogate safety indicator. Note that a column for outcomes has been added since it is also possible for surrogate indicators to measure based on the outcomes of a traffic event.

Table 2. Summary of Surrogate safety indicators and their relation to collision risk and injury risk. Taken from (Johnsson et al., 2018).

Indicators	Collision risk			Injury risk	No. of validation studies	
	Initial conditions	Evasive actions	Outcomes		All	Including VRUs
TTC based	No	No	Yes	No	3	2
Time to Accident	Yes	No	No	No	2	2
PET based	No	No	Yes	No	7	1
Conflict Index	No	No	Yes	Yes	1	0
Deceleration based	No	Yes	No	No	0	0
Yaw Rate	No	Yes, for motorcyclists	No	No	0	0
Pedestrian Risk Index	Yes	Yes, for motor vehicles	-	Yes	0	0
Conflict Severity	Yes	No	-	Yes	0	0
DeltaV	No	No	No	Yes	0	0
Extended DeltaV	No	No	Yes	Yes	0	0
Change in Step Frequency	No	Yes, for pedestrians	No	No	0	0
Aggregated Crash Index	Yes	Yes	-	No	0	0
Space Occupancy Index	Yes	No	No	No	0	0
The American TCT	No	Yes	No	No	4	0
The Canadian TCT	No	Yes*	Yes	No	3	3
The Dutch TCT	Yes*	Yes*	Yes	Yes*	2	2
The Swedish TCT	Yes	No	-	No	3	3
The British TCT	Yes*	Yes*	-	Yes*	3	0

**The asterisk indicates that the indicator relies on a subjective component.*

Only a few indicators perform well from the perspective of collision risk and injury risk. Both of the most used indicator types (TTC_{min} and PET) instead identify critical events based on the outcome of a traffic event. That is, the severity according to these indicators depend both on the initial conditions and any evasive actions taken during the event, but the indicators make no attempt at separating them. Estimating the severity of a traffic event based on the outcome of the event can potentially have some problems. It is, for instance, possible that an event with severe initial conditions, followed by an effective evasive action, can lead to a non-severe outcome, and therefore not be identified as critical. For example, a motor vehicle braking in front of a pedestrian can create a high PET value (indicating low severity), even though the situation was severe. Another example might be a swerving cyclist who can quickly remove him/herself from a collision course with a motor vehicle, before the situation would become critical according to TTC_{min}.

There are some indicators that estimate the severity of traffic events based solely on the magnitude of evasive actions. These indicators frequently focus on identifying severe braking, but there are also some indicators that focus on swerving and running (Tageldin & Sayed, 2016; Tageldin et al., 2015). Relying solely on evasive actions can also result in some potential problems. It is possible to observe braking, swerving, stopping, or running without severe initial conditions. For example, a running pedestrian is not necessarily avoiding a collision, but could instead be in a hurry or running to quickly allow the motor vehicles to pass. Relying on evasive actions also disregards situations without evasive actions which might still be severe. For example, it is possible for two road users to be very close to colliding without anyone of them reacting or even seeing the other.

Relatively few indicators estimate the initial condition in a severe situation. The TA indicator measures the time to collision at the start of an evasive action. While not strictly measuring the initial conditions of a situation, TA does estimate the conditions for evasive actions, which is similar to the function of the initial conditions described by Davis et al. (2011). Both the Swedish TCT and the Conflict Severity Indicator (Bagdadi, 2013) rely on TA, but both combine TA with speed and the road user's assumed deceleration to estimate the severity of the initial conditions. However, neither of them considers any other type of evasive action. A potential solution to this problem is to limit the type of traffic situations in which an indicator can be applied. Both the ACI and the PRI (Cafiso et al., 2011) rely solely on deceleration, but they are only useable in very specific situations. However, note that the PRI does not consider any evasive actions taken by the pedestrians.

Finally, some of the indicators (the PRI, the Dutch and British TCT) manage to include all three of the aspects discussed in this review: the initial conditions, the evasive actions, and risk of injury. However, these indicators either limit themselves to a very specific situation, such as the PRI, or rely on the subjective evaluation of the severity by human observers, such as the Dutch and British TCT.

Validity

The extent of validity of the different indicators is difficult to appraise due to the widely different methods used and the equally large divide in the number of observation locations and duration of observation. However, almost all the validation studies found in this review do show a significant correlation between number of critical events and crashes regardless of what indicator is being used. The result also shows that while many validation studies have solely focused on motor vehicles, some studies have included VRUs. However, not all of these studies have done a separate analysis for VRUs, and only one study (Lord, 1996) has focused on pedestrians. No validation study has specially focused on cyclists.

The findings show various approaches used to validate surrogate indicators, and it is unclear whether any particular method is preferable. The number of studied locations and the duration of the study at each location also fluctuate greatly among the studies, which further complicates any attempt to compare the extent of validation among the different indicators.

Another question, which is discussed by some of the studies and further discussed by Güttinger (1982), is how strong the connection between critical events and crashes must be to be considered good enough. The report by Migletz et al. (1985) provides a potential answer by remarking that the critical events are good estimates of safety, since they produce estimates of the expected number of crashes better or equally as good as historic crash data. Güttinger (1982) also suggests a similar answer in that the relation between critical events and crashes must be stronger than the relation between exposure variables.

Universal and specific indicators

Certain indicators (such as TTC-, PET- and DeltaV-based indicators) aim to provide a universal estimation of severity regardless of the involved road users or the specific setting. Most of the indicators of the first (universal) kind focus on cars. Indicators such as the Swedish TCT, the Conflict Severity, the Extended DeltaV, and the different deceleration-based types include assumptions regarding the deceleration of cars. Even TTC- and PET-based indicators concentrate on cars, since the commonly used threshold values are chosen based on the observation of cars (Hayward, 1971; van der Horst, 1991). Other indicators are designed to only be used in very specific settings, while others are meant to be more universal. The ACI (Kuang et al., 2015), PRI (Cafiso et al., 2011), and the Space Occupancy Index

(Ogawa, 2007) are examples of indicators meant to be used only in very specific settings such as on motorways or on protected bikeway paths.

Both universal and specific indicators have their advantages and disadvantages. Universal indicators can be applied in a variety of different types of settings, which makes them very useful. However, specific indicators can more easily consider all important factors in their specific setting, which allows for a more comprehensive indicator within that setting.

Reliability

The most obvious reliability concern is the use of subjective judgement. Indicators based on human judgement can potentially allow for the consideration of both the initial conditions, the potential evasive actions, and risk of injury in more detail than any other indicators. However, the use of a subjective component also has several drawbacks, especially the potential reliability concerns and the difficulty of incorporating such components into automated systems. The use of subjective components and trained observers has historically been common in TCTs.

Some of the literature indicate the possibility of achieving satisfactory levels of reliability among different observers (Grayson, 1984; Kruysse, 1991; Kruysse & Wijlhuizen, 1992; Shinar, 1984). The so-called Malmö Calibration Study (Grayson, 1984) also provides a good comparison between several TCTs and objectively measured values using computer vision. The comparison showed that teams applying different TCTs agreed well on their allocation of critical events regardless of road user type.

Motion prediction

In addition to human observers and subjective assessment it is also possible to discuss reliability between different computer-based systems. This is especially relevant for trajectory prediction, which is necessary to identify a collision course. Predicting the future path and speed of a road user can be done using several methods – from a simple approach which relies on constant speed and assumptions about how the road user intends to travel, to more complex methods based on observed behaviour at the studied locations (St-Aubin et al., 2014). The choice of method for predicting the path of road users will influence what situations are identified as severe, regardless of which indicator is used. This makes it further complicated to compare surrogate safety indicators, even if they are based on objective values such as TTC_{min}. This is particularly relevant whenever swerving is used as an evasive action, since this makes it difficult to accurately predict the path of the road user.

Conclusions

There are many different surrogate safety indicators suggested in literature. The review shows that while many indicators have been suggested in the literature, TTC_{min} and PET are most commonly used. However, few of the indicators focus on aspects that are important when studying VRUs. The results show that many indicators focus on braking as an indicator to identify critical events and does not consider swerving or running. Swerving also creates concerns for the indicators that rely on collision course, since predicting the path of a swerving cyclist seems very difficult compared to predicting the path of a braking motor vehicle.

Few indicators try to estimate the injury risk. The large difference in mass and the lack of a protective shell for the VRUs makes the risk of injury very different for events involving only motor vehicles compared to events which involve VRUs. Overall, very few indicators manage to consider all the relevant aspects of importance, and those who do either use subjective evaluation, such as the Dutch and British TCTs, or are limited to a very specific setting, such as the PRI. However, it should be noted that while the PRI does technically consider all aspects, it does not consider any evasive actions from the pedestrians which might be relevant.

It is noteworthy that even though no single indicator reflects all relevant aspects discussed in this analysis, all aspects are accounted for by the different indicators. Deceleration, Jerk, Yaw Rate, and Change in Step Frequency all attempt to capture evasive actions from different types of road users. Time to Accident, PRI, and Space Occupancy Index attempt to identify severe initial conditions, and Delta-V attempts to evaluate the injury risk dependent on the types of road users involved in a critical traffic event. Further research focusing on combining the many different indicators could help to strengthen the theoretical base for SMoS.

Finally, the results show that there are some validation studies that have included VRUs, but the many different approaches used, the length of observations, and the number of locations studied in the validation studies makes comparing the extent of validity between the indicators difficult. However, there seems to be a consistent and significant correlation between numbers of critical events and crashes regardless of which indicators are used, and comparisons between TCTs seem to indicate significant similarities between the results from different indicators. Further research into the validity of surrogate safety indicators is needed. The research should also attempt to answer what extent of validation is good enough and to what extent validation of one indicator can be used to evaluate other indicators.

PART 2

4. Observational studies

The aim of this part of the thesis is to attempt to answer the research questions posed using the findings from the observational studies presented in paper 2, 3, 4, and 5. This part of the thesis mostly builds on the video data gathered in the European project *InDeV – In-Depth Investigation of Accident Causation for Vulnerable Road Users*. The following chapters describe the data collection and processing, followed by the main attempt at validation. The results and conclusions from that attempt then function as the basis and motivation for the remaining chapters.

5. Data collection and processing

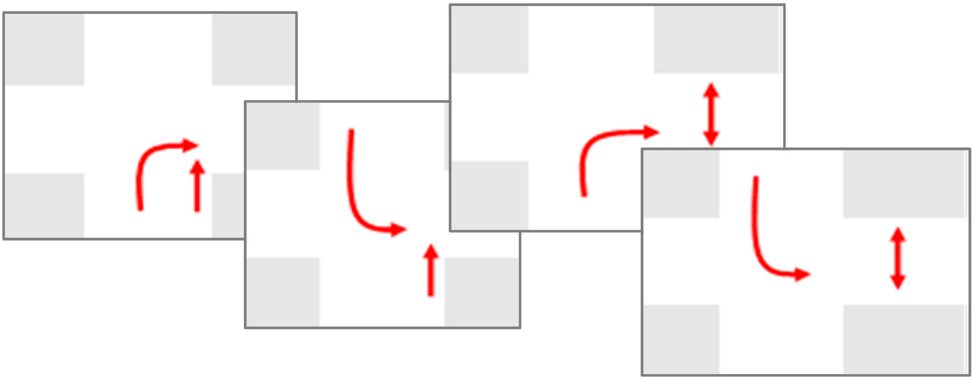
This chapter presents data gathered within the InDeV project. The project covered 26 sites in 7 European countries, studied for 3 weeks. This data functions as the base for the work presented in the following chapters of the thesis, however, some additional data was gathered from the same video recordings as described in chapters 8 and 9.

The main aim of the data described in the following chapter is to perform a large-scale attempt at validating SMoS applied to vulnerable road users (described further in chapter 6). Both crash- and SMoS data was therefore gathered. SMoS data collection was further split in two parallel tracks since *severe* events are rare and require long observation periods, while the *normal* events occur frequently and thus can be collected within short time.

Crash types of interest and site selection

The following criteria for crash types were set: *1*) the crash must involve a motor vehicle and a vulnerable road user (pedestrian or cyclist); *2)* the crash must result in a fatality or injury; *3)* the crash must occur at an intersection. The first criteria was imposed by the nature of the InDeV project that had a clear focus on vulnerable road users; property-damage-only crashes were excluded since their reporting is down-prioritized in most of the European countries in favour of injury crashes, thus the figures are either unavailable or unreliable; intersections were chosen since SMoS data was collected through video-filming with stationary cameras, thus events should be *concentrated* in space to be repeatedly captured by the camera.

The latest version of the data structure for the European Union Crash Database CARE (CaDaS, 2015) contains a crash typology (a set of codes with corresponding sketches) that should have made finding the most frequent crash types fulfilling all the criteria a trivial task. However, this part of the CARE is still mostly empty as only two countries – Germany and Denmark – use similar systems at the national level. Thus, the decision was made based on data from these two countries and complemented by the results of manually processed crash records from two large Swedish cities (Björnberg, 2016). The selected crash types are shown in Figure 13. It is important to note that these types are not the most frequent among all pedestrian/cyclist crashes, but only among those that fulfilled the stated criteria (the absolute leaders are single crashes standing for more than 60% of all severe injuries - Björnberg (2016)).



a) *b)* *c)* *d)*

Figure 13. Most frequent crash types selected for further analysis: a, b) motor vehicle right/left - cyclist straight; c, d) motor vehicle right/left – pedestrian crossing the intersection approach.

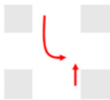
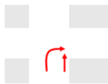
In total, 26 intersections were selected for analysis in seven European countries (Belgium, Denmark, the Netherlands, Norway, Poland, Spain, and Sweden). Despite great variety in how bicycle/pedestrian facilities are designed in these countries, a significant effort was made to find comparable locations. All sites are signalized intersections having simultaneous green for involved motor vehicles and cyclists/pedestrians' directions, no separated phase or pre-/after-green for turning manoeuvres. Within each country, additional criteria were set for consistency in the design among the sites; for example, if it is a bicycle on-road lane or a separated track adjacent to the pedestrian crossing. In Spain, all the studied locations were on one-directional roads, thus no left turning manoeuvres are possible.

Crash data

The crash history (period 2009-2016, with minor deviations for some countries) from the selected sites revealed that the data was not enough for estimation of the expected crash numbers. Once disaggregated based on the manoeuvre, most of the data cells contain one or zero. Therefore, it was decided that crash models based on a larger number of similar sites would be developed. However, even when the dataset is extended to ca 50 similar locations per country, the crash numbers are still low (Table 3).

To control for the exposure, traffic flow data was collected. The Annual Average Daily Traffic (AADT) for each manoeuvre and road user category was estimated based on 45 minute-counts on a weekday in spring/autumn and a daily profile for a specific road user category. Some countries had average daily profiles recommended by the authorities, while in other countries the profile shapes were estimated based on manual counts from the available video footage.

Table 3. Disaggregated injury crashes records per country*.

	Number of sites				
Belgium	*50*	*2*	*2*	*4*	*4*
Denmark	*50*	*5*	*8*	*2*	*1*
Norway	*79*	*6*	*15*	*0*	*6*
Poland	*50*	*1*	*1*	*2*	*11*
Spain	*27*	*28*	-	*3*	-
Sweden	*36*	*24*	*28*	*1*	*3*

* *The crash data from the Netherlands was practically unavailable, and is therefore not included here.*

SMoS data

Collection method

All 26 locations were filmed during at least 3 weeks. The video data was collected using three cameras (one thermal and two RGB), primarily for the purpose of evaluating the camera perspective and sensor type effects on the video processing tools. At the end, the video with the best view of the studied manoeuvres was used.

The SMoS data collection was split in two parallel tracks since *severe* events are rare and require long observation periods, while the *normal* events occur frequently and thus can be collected within short time. This split resulted in two datasets, the first containing all meetings from a 1-day period, and the second containing chosen events from the full 3-week period. Furthermore, all SMoS data collection was made in a two-step process in which a human observer first identified a relevant situation, and then trajectories for this situation were manually created.

1-day data

The aim of the 1-day data collection was to capture all meetings for 24 hours. After some early testing, it became apparent that a meeting (i.e. a simultaneous presence of two road users heading towards a common conflict area) does not always result in a clear interaction. Therefore, a set of additional operational rules were used to only include meetings with a more *direct interaction*. Firstly, each meeting only included one motor vehicle and one vulnerable road user, meaning that even if there were multiple VRUs passing in front of the motor vehicle, only one of the VRUs was chosen (the choice was made by selecting the VRU that was closest to the motor vehicle while it was still in motion). Secondly, a meeting was excluded if any one of the road users was standing still during the entire passing of the other road user, i.e. only situations in which both road users moved were included.

Note that the identified events based on these rules will henceforth be called *encounters*. However, three other definitions of encounters are also defined and further explored in chapter 8.

For a 24-hour period, all encounters were detected manually, and trajectories of the involved road users produced. This was done using the semi-automated tool T-Analyst (T-Analyst, 2019) which allows managing large amounts of video data and making bookmarks (detections) in it (Figure 14). It should be noted that in T-Analyst, any calculations that require motion prediction (such as calculating TTC and T2) are made based on the assumption that the road user will travel along the actually revealed trajectory, but keep the constant speed as at the moment of calculation (detailed procedure can be found in A. Laureshyn et al. (2010)).

Figure 14. Screenshot of T-Analyst software (T-Analyst, 2019).

3-week data

For most of the sites, a 24-hour period provided several hundred individual traffic events, which made a solid basis for analysing the distribution of SMoS in normal (non-critical) traffic conditions. However, the number of critical events, which are most relevant in SMoS context, was expected to be low in this dataset. To complement this part of the SMoS distribution with more data, only critical incidents were selected in the remaining video (ca 3 weeks).

Detection of rare events in 3 weeks of video is a demanding task. It was partly automated by using a watchdog software called RUBA (Madsen et al., 2016). The basic functional unit of RUBA is a detector (a certain area of the image monitored for changes), which can detect presence, motion in general, or motion in a certain direction. By strategically placing the detectors and defining the rules for temporal relations between the activations of the detectors, it is possible to find the *simultaneous presence* of two road users. The observer can thus focus on these parts of the video, and skip watching the parts where nothing relevant occurs.

The limitation and main drawback of the watchdog tool is that while it can detect a potentially relevant event, it does not provide any indication of how severe the event is (and thus cannot help the observers to focus on the most severe ones). The information about the time between the detector activations is not of much use when the speeds of the road users are unknown. After several attempts of adding further *intelligent selection* steps, and comparison of the results with the manually produced data from the 24-hour period, it became clear that the observers have to watch all the detections to ensure that no important events go undetected.

How the observers select the situations has an impact on the results, therefore feasible precautions were taken. The observers were instructed to select all out-of-the-ordinary situations that in some way were perceived as risky, dangerous, or out-of-control. In case of doubt, the situation was to be included. After the relevant situations had been found, trajectories of the involved road users were produced in T-Analyst. The severity of these events could then be estimated by an objective indicator (such as TTC or PET), which would, in theory, reveal all critical events.

1-day data description

This section gives an overview of the one-day dataset generated from the study sites. Note that the dataset include data from only 21 locations. Due to time limitations within the project, the work was not completed at three intersections in Spain, one in Belgium, and one in Poland. Table 4 shows the number of identified encounters divided into the four relevant manoeuvre types. The dataset includes an even split between cyclists and pedestrians, with approximately 4500 encounters processed for each type of road user, but the right-turning manoeuvres outnumber the left-turning manoeuvres for both cyclists and pedestrians.

Table 4. The number of encounters for each manoeuvre processed within the 1-day dataset.

	Number of sites				
Belgium	*3*	*481*	*105*	*374*	*96*
Denmark	*4*	*458*	*445*	*200*	*169*
Netherlands	*4*	*987*	*135*	*335*	*40*
Norway	*3*	*282*	*77*	*727*	*174*
Poland	*3*	*61*	*37*	*1013*	*399*
Spain	*1*	*421*	-	*386*	-
Sweden	*3*	*769*	*347*	*412*	*199*
Total	*21*	*3459*	*1146*	*3447*	*1077*

3-week data description

This section gives an overview of the 3-week dataset. The data includes only 15 locations. In addition to the locations missing from the 24-hour dataset, the 3-week dataset is missing all of the locations in the Netherlands and one of the locations in Poland. Also note that additional effort went into the single Spanish location, however this effort did only focus on events including bicyclists.

Table 5 below shows the number of severe traffic events identified by the human observers based on the 3 weeks of video data from each of the intersections, separated both by the different manoeuvres and by the type of road user. The table also shows how the 355 days of observation are divided between the different countries (approximately 23 days of observation at each site). Note that the large number of pedestrian events selected in Poland were caused by a misunderstanding that resulted in many events considered only *slightly severe* being included in the dataset.

Table 5. The number of events for each manoeuvre processed within the 3-week dataset.

	Number of sites/ Days observed				
Belgium	3/59	161	52	34	42
Denmark	4/91	177	178	96	103
Netherlands	0	-	-	-	-
Norway	3/63	149	63	68	41
Poland	2/42	52	153	562	681
Spain	1/41	142	-	-	-
Sweden	3/59	39	31	17	10
Total	16/355	559	477	777	877

6. An attempt at validation of SMoS

This chapter presents the main findings from a large-scale attempt to validate SMoS focused on vulnerable road users. The general approach to the validation is illustrated in Figure 15. The chosen approach has two parallel problems that require solutions: 1) detection of the relevant safety critical events; 2) estimation of the expected number of crashes as the direct measure of safety at a studied site (the *ground truth* to compare with SMoS).

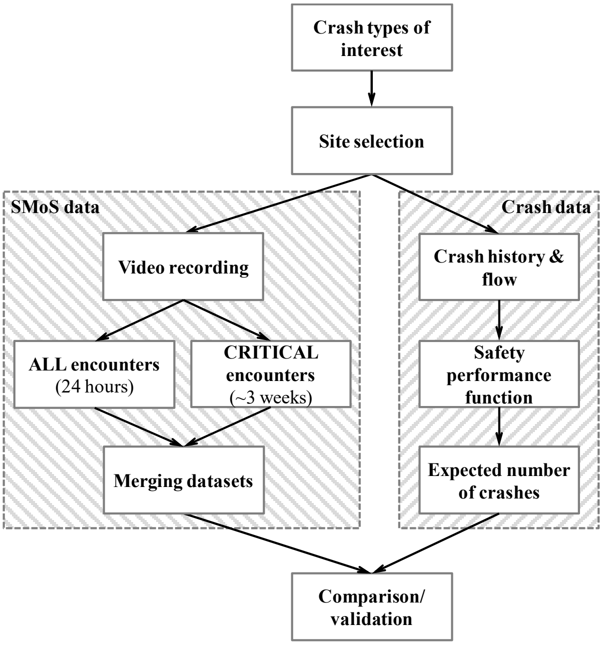


Figure 15. The general validation algorithm.

Considering the various different approaches to validation found in literature (page 14), the chosen approach is similar to the one used by Lord (1996). This method was chosen due to the relatively long observation period (3 weeks) at each location, which limited the number of sites that could be studied.

The limited number of locations together with a low number of recorded crashes at each site made it impractical to directly study the relation between critical events and recorded crashes. By instead estimating the expected number of crashes at each site using a safety performance function (or crash model), it was possible to gain a more robust estimate of the safety from each of the observed locations. To make the crash estimate as robust as possible, crash records and traffic counts from various similar intersections were gathered in each country (Table 3).

However, analysis of the data exhibited two significant shortcomings which made the original approach to testing the validity infeasible. These shortcomings were:

1. The difficulties of comparing crash data from different European countries
2. The problem of discrepancies when merging the SMoS datasets

These shortcomings limited the scope of the validation, with the result being a study of 9 Scandinavian intersections with a focus on cyclist-motor vehicle interactions. The following sections will further describe both shortcomings in more detail, followed by the downscaled validation study based on the Scandinavian data.

Crash data from different European countries

Looking at the gathered crash data (Table 3), there are some immediate causes for concern. How is it possible that Sweden, one of the best performing countries within the domain of road safety, has such high crash numbers compared to other countries? Is there a non-negligible under-reporting that is not balanced among the countries?

The bias introduced by the under-reporting becomes much more evident when a crash model is calibrated using the crash and traffic data for each country. Consistent with the state-of-the-art in crash modelling (Lord & Mannering, 2010; Mannering & Bhat, 2014), the negative binomial model form was assumed:

$$E(y) = e^{a_0} \cdot ADT_{Veh}^{a} \cdot ADT_{VRU}^{b} \cdot e^{Country} \qquad (1)$$

where *E(y)* is the predicted yearly crash frequency, ADT_{Veh} and ADT_{VRU} are the traffic flow values for motor vehicles and VRUs respectively, a_0, a, and b are regression parameters to be estimated, and *Country* is a categorical variable with a value for each single country. A range of models have been tested varying in the levels of input data disaggregation (left/right turn vs. bicycles/pedestrians, motor vehicles vs. bicycles/pedestrians, motor vehicles vs. VRU, etc.), all suffering from the same type of issues. To save space, only one of the model results are provided as illustration in Table 6.

Table 6. Crash prediction model for right-turning motor vehicles and cyclists (crash type a in Figure 13).

Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		P-value	$e^{Country}$
a_0	-7.79	3.55	-14.75	-0.84	0.0281	
a	0.63	0.36	-0.08	1.33	0.0828	
b	0.31	0.25	-0.19	0.80	0.2284	
Country (BE)	-2.58	0.93	-4.40	-0.76	0.0054	0.08
Country (DK)	-2.87	0.86	-4.56	-1.18	0.0009	0.06
Country (NO)	-0.62	0.70	-2.00	0.75	0.3742	0.54
Country (PL)	-1.66	1.08	-3.77	0.45	0.124	0.19
Country (SE)	0.00	0.00	0.00	0.00	-	1.00
Dispersion	1.08	0.95	0.19	6.02		

From an analytical point of view, the main problem is that no statistically significant relationship can be found between crash frequency and exposure. Moreover, the values of $e^{Country}$, though in most cases not statistically significant, indicate that all countries are systematically and dramatically safer than Sweden. This pattern is consistent for all crash types and aggregation levels. There are two likely causes for this problem: 1) there are too few crashes within the data, and 2) there is likely a significant amount of under-reporting within the data.

The issue of under-reporting is not easily mitigated. The most common measure suggested to reduce under-reporting is to link police data with the injury records from hospitals and other medical institutions (Elvik & Mysen, 1999; Yannis et al., 2014). The true scale of the problem can be estimated through self-reporting methods that can *reach* even injuries registered neither by police nor hospitals (Andersen et al., 2016). Even though there is a handful of studies performed in different countries attempting to quantify the under-reporting rates for different crash severity levels and road user types involved (Amoros et al., 2006; Elvik & Mysen, 1999; Janstrup et al., 2016; Kamaluddin et al., 2019), these findings are hard to use for improvement of the developed model. Usually, the results are presented on an aggregated level of countries or regions, and do not distinguish crash location, manoeuvre type, etc. Another relevant issue, pointed out by Olszewski et al. (2016), is the lack or inconsistency of definitions for injury crashes among EU countries, meaning that different practices exist for which crashes are being reported or not.

Scandinavian crash model

Attempts to develop crash models based on only Scandinavian (Denmark, Norway, and Sweden) data were also made, the idea being that these countries have a similar traffic culture and might therefore be more comparable. The resulting model (Table 7) includes crashes between cyclists and turning motor vehicles. CURE plots (Hauer & Bamfo, 1997) indicated a good model fit, as cumulative residuals do not exceed the boundaries. Note that the pedestrian crash data was excluded since it resulted in P-values higher than 0.05, likely due to the even lower number of recorded pedestrian crashes (Table 3).

Table 7. Crash prediction model for turning motor vehicles and cyclists in Scandinavia.

Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		P-value
a_0	-10.72	2.78	-16.17	-5.27	0.0001
a	0.80	0.28	0.26	1.35	0.0041
b	0.60	0.20	0.21	0.99	0.0026
Country (DK)	-1.71	0.49	-2.66	-0.75	0.0005
Country (NO)	-0.12	0.53	-1.16	0.92	0.827
Country (SE)	0.00	-	-	-	-
Dispersion	0.55	0.43	0.12	2.54	

Discrepancies in the SMoS data

As described in Figure 15, the SMoS data was collected in two parallel tracks. For the 24-hour period, trajectories for all relevant encounters were created; for the remaining 3-week period, among all detected events only those that seem to have some degree of dangerousness were further processed. Obviously, while 24-hours provide a good description of *normal traffic*, the number of severe situations that can be captured is low, and their number is greatly affected by randomness. On the other hand, the 3-week period contains a more solid collection of severe events, but as the severity decreases, the events appear in the *grey zone* of inclusion/non-inclusion by the observer, and their frequency cannot be trusted any more.

Based on this idea, the hypothesis was that it is possible to find a convergence point beyond which the results are consistent, i.e. at a certain threshold it should be preferable to stop relying on the 24-hour data, and instead start to rely on the 3-week data, since the 3-week dataset should contain a better estimate of the frequency of severe events. However, the data does not show any such breaking point. Instead, the 24-hour data consistently show a higher daily frequency of critical events compared to the 3-week data, regardless of which indicator was tested or their corresponding threshold values.

Table 8 presents estimated daily frequencies of events in different severity categories defined by two indicators – TTC_{min} and PET. The data is aggregated for 15 intersections (all intersections which had a completed 3-week analysis including both bicyclists and pedestrians), which results in a total of 355 days for the 3-week datasets and 15 days for the 24-hour datasets.

Table 8. Observed daily frequencies of events in different severity categories.

Daily values					
	3-weeks	24-hours		3-weeks	24-hours
TTC_{min}<5s	1.7	63.3	PET<5s	3.1	184.1
TTC_{min}<4s	1.7	55.1	PET<4s	3	178.8
TTC_{min}<3s	1.6	35.5	PET<3s	2.9	166.1
TTC_{min}<2s	1.3	9.9	PET<2s	2.7	124.9
TTC_{min}<1.5s	0.7	3.7	PET<1.5s	2.2	80.7
TTC_{min}<1s	0.2	1.4	PET<1s	1.3	31

Obviously, the two datasets never converge. Even in high-severity categories, the observers systematically select fewer events compared to the strictly objective selection of the 24-hour dataset. The hypothesis was tested that conservatism of human observers might lead to underreporting the encounters. To investigate this issue, the events from the 24-hour dataset in categories of TTC_{min} <1.5s and PET<1.5s were watched and discussed with experts in SMoS. The conclusion was that the problem was caused by the ineffective indicators used, which were not very good at reflecting what a human would perceive as dangerous. While the situations with low TTC or PET included events that seemed to be break-downs, dangerous, and out-of-control, they included a much greater number of situations that were in perfect control by the involved road users that hardly appeared to hold any risk of collision, not to mention injuries.

Scandinavian validation study

Following the limitations set by the two shortcomings described above, a choice was made to make a downscaled validation study using only Scandinavian data, excluding the 3-week data from the analysis and focusing solely on the 24-hour dataset. Note that one of the Danish intersections were excluded, since there were no left-turning motor vehicles at that site.

Starting with the crash data, the predicted number of crashes at each of the Scandinavian locations was calculated using the Scandinavian model presented in Table 7. Following that, the expected number of crashes at each of the studied locations was estimated using the Empirical Bayes Correction (Hauer, 1997b). The correction combines the predicted crash number from the model with the number of recorded crashes based on the dispersion of the crash model. This allows the resulting estimate of the expected number of crashes to consider both the underlying relations established in the model and also consider the local conditions at each location using the number of recorded crashes.

The resulting crash estimates, the number of encounters, and the corresponding number of critical events using both TTC_{min} and PET can be seen in Table 9. The table also shows the R-squared (R^2) value, indicating the linear correlation between the expected number of crashes and the SMoS indicators at different threshold levels, and the correlation between the SMoS indicators and the encounters. Note that the column $TTC_{min} < \infty$ indicates the total number of encounters that had a collision course at some point, regardless of the TTC value.

The result from the SMoS analysis shows that, at least for some threshold values, there is a substantial correlation between the expected number of crashes and the number of observed critical events. For these thresholds, the results are comparable to what has been published in earlier validation studies. However, there are three major concerns. Firstly, contrary to the SMoS theory, the correlations do not improve but rather dramatically deteriorate as the thresholds for TTC_{min} and PET are set lower. For TTC_{min}, this can be partly attributed to the low number of events selected by a low threshold, leading to a higher sensitivity to random variation.

Table 9. Daily number of encounters, critical events using different threshold values, and the expected number of crashes per year.

Site	Estimated Crashes/ year	Daily values									
		ENC	TTCmin					PET			
			< ∞	< 4	< 3	< 2	< 1.5	< 5	< 3	< 2	<1
DK 1	0.11	51.3	7	5.8	4.5	1.8	0.8	48.8	47.3	39.3	13.5
DK 2	0.15	179	63	45	23	5	3	179	176	165	64
DK 3	0.22	102.5	8.5	5	3.5	1.5	1	97.5	97.5	89.5	27
NO 1	0.49	116	28	25	16	5	3	108	85	64	20
NO 2	0.12	155	56	45	31	11	2	147	129	100	33
NO 3	0.69	117	38	28	17	5	3	112	105	90	28
SE 1	3.14	310	96	83	53	10	1	258	231	135	13
SE 2	2.79	345	201	144	85	12	2	330	307	228	38
SE 3	0.11	369	74	65	35	10	1	312	267	160	21
R^2, related to encounters		-	**0.80**	**0.86**	**0.83**	**0.82**	**-0.21**	**0.99**	**0.98**	**0.84**	**0.04**
R^2, related to crash estimate		**0.73**	**0.62**	**0.64**	**0.58**	**0.49**	**0.05**	**0.76**	**0.73**	**0.59**	**0.00**

Secondly, there is significant correlation between encounters and both TTC_{min} and PET, even at relatively strict threshold values. As expected, a more lenient threshold makes the correlation stronger, while a stricter threshold makes the correlation less pronounced. Finally, neither TTC_{min} nor PET shows a stronger correlation with crashes compared to the encounters. Following the idea that SMoS are meant to function as a surrogate to crashes, i.e. they should be dependent on both exposure and risk, it would have been expected that the indicators would outperform measures that only attempt to measure exposure.

Considering these points, the result looks quite discouraging for the validity of SMoS. Even though there is a substantial correlation between the number of critical events and crashes, the result suggests this correlation principally originates from the strong relation between encounters and crashes and the inherent connection between critical events and encounters. However, it is also possible to study the number of critical events that occur per encounter at the different intersections. This value represents the risk at each site and can be compared with the result from the Scandinavian model (Table 7).

The result from the crash model suggests that the Danish intersections are generally safer compared to the Norwegian and Swedish sites which are themselves about equally safe. Looking at the average number of critical events per encounter from the SMoS study (Table 10), the same result is found using TTC_{min} with a threshold higher than, or equal to, three seconds, while PET continually disagrees with the crash model, regardless of threshold value. This result suggests that while TTC_{min} fails to outperform encounters on an individual site level, it might still hold some additional information about risk when considering several intersections together.

Table 10. The average number of critical events per encounter in each country.

Country	**$TTC_{min} < \infty$**	**TTC_{min}<4 s**	**TTC_{min}<3 s**	**TTC_{min}<2 s**	**TTC_{min}<1.5 s**	**PET<5 s**	**PET<3 s**	**PET<2 s**	**PET<1 s**
DK	0.24	0.17	0.09	0.02	0.01	0.98	0.96	0.88	0.31
NO	0.31	0.25	0.16	0.05	0.02	0.95	0.82	0.65	0.21
SE	0.36	0.29	0.17	0.03	0.00	0.88	0.79	0.51	0.07

Discussion

The original aim of the study was to validate the use of various SMoS with a specific focus on VRUs. However, the two problems discussed above hindered a large-scale validation study, resulting in a downscaled study focusing solely on Scandinavian data. There are also several noteworthy aspects to consider for future studies.

The first problem of limited crash data is a major challenge. The lack of crash data makes any attempt at creating disaggregated crash models very resource intensive. The differences in under-reporting rates do not allow for building a cross-country crash model either. While fatal crashes could be expected to be reported reliably in most countries, they are so few that the number of sites necessary for building a model becomes unrealistic (within the Swedish dataset, only 2 crashes were fatal). This problem has been found in early validation studies (Migletz et al., 1985; Å. Svensson, 1992), and it seems to become even more acute in modern times. On the other hand, this is a strong argument for further development and use of the SMoS, as crashes simply cannot be used for measuring safety unless aggregated on a high level.

The second problem of discrepancies found while merging the 1-day and 3-week data points out several issues, too. First, neither TTC nor PET, both very commonly used indicators, seem to reflect severity as it is perceived by a human. As a result, the severity rankings based on TTC or PET seem counter-intuitive when individual events are actually examined. While the events judged as *severe* by humans do indeed have low TTC or PET, the opposite is not true. There is no obvious proof that human perception of severity is a reflection of the *true* severity dimension, however, it is clearly more comprehensive, and covers aspects such as nearness-to-collision, consequences-if-collision, level-of-control during the situation, etc. Earlier studies found strong agreement among humans in ranking the situations by their severity (Asmussen, 1984; Kruysse, 1991), indicating that there is some universal instrument for judging risks (at least when observing a situation as a third party).

Clearly, there is a large potential for improvement here, and a need for indicators that are more comprehensive, taking different aspects of a situation into account.

It should also be noted that calculation procedures for the indicators seem to contain certain challenges. TTC, the indicator most frequently used in SMoS studies, was calculable in only 35% of all situations, making the rest of the data unusable. Indeed, in some situations seeming to have a collision course, the road users are in fact separated by a tiny time gap, which becomes apparent when TTC is calculated using correct dimensions of the road users and accurate and realistic trajectories.

This again makes the selection different from what was done by human observers in earlier studies, as such situations were included and a TTC estimate was produced for them, too. Another dimension of the problem is unrealistic assumptions in calculations: for example, that the speed will remain constant during the entire manoeuvre. Therefore, more advanced methods for future motion prediction might be necessary as, for example, discussed in Mohamed and Saunier (2013). However, it should also be noted that we can expect very small TTC estimates to be more robust since there is less time for a road user to make alterations to their trajectory.

Finally, the validation study made using the Scandinavian data showed several noteworthy results. The study found a significant correlation between crashes and both TTC_{min} and PET, however, the result also suggests that this might be due to the strong connection between encounters and the resulting critical events. This suggests that the relation between SMoS and elementary units of exposure should be further considered in future research. It is important that SMoS have stronger relation to risk than the exposure measures (Güttinger, 1982; Hauer, 1982), to provide *additional value* (and justify the additional costs related to the SMoS collection). As has been shown above, lenient SMoS thresholds select events that are highly correlated with exposure measures and thus might not really contain any additional information about risks. The property of *not being the same as exposure* can thus be used as an indirect criterion that the SMoS is *behaving* properly. Obviously, it is also important that any correlation between SMoS and crashes does not originate in their inherent connection to encounters.

7. A relative approach to validation

The complications discussed in the previous chapter make classical validation difficult and resource intensive. The largest problem is the lack of crash data when disaggregating based on different manoeuvres and types of design. The strong connection between encounters and SMoS should also be considered, since it is crucial that any usable SMoS cannot simply be replaced by encounters.

This chapter aims to discuss relative validity as a different approach to investigating the validity of SMoS. This indirect approach to validity could allow for less resource intensive studies, which could be useful in identifying poorly functioning SMoS, and providing a further step if classical approaches to validity are infeasible. The approach also directly considers the link between encounters and SMoS in a practical way.

Concept of relative validity

Relative validity refers to a less strict, but possibly more practically feasible, approach to the validity of SMoS. The core idea is that while a certain SMoS might not estimate the expected crash number, it still works well in making a comparison between different sites or conditions at the same site. Basically, a relatively valid SMoS is able to reflect the direction of the change in safety (improvement or deterioration), but unable to describe its magnitude. Despite this limitation, the result of such a study might still provide valuable information.

The use of relative validity removes some of the practical issues, compared to a more classical approach. With a relative validation study, comparison of only two sites already contribute useful pieces of knowledge. Such small-scale studies are easier to perform, and therefore one could expect that the knowledge accumulation would go faster, though with smaller steps. Another implication is the position of the threshold between *near-crashes* and *normal events*. Theoretically, if the goal is to estimate the expected crash number, the strictest possible threshold is to be preferred, as those events should be the *closest to crash*. This makes the frequency of the relevant events low and the necessary observation periods long. For the relatively valid SMoS, the threshold should be merely high enough to distinguish between safety of two sites/conditions and, presumably, will be much more inclusive. In practice, again, that would mean that both validation studies – and practical safety assessments based on SMoS – could be performed during a shorter period of time and thus more of such studies could be expected.

An outline for testing the relative validity of a SMoS

Using the concept of relative validity and the idea of elementary units of exposure, it is possible to construct a practical approach to testing the validity of a SMoS. Such an approach would consist of two main parts: the *ground truth* and the SMoS diagnosis.

The *ground truth* refers to a known safety ranking between two different types of infrastructure designs. The main aim of the ground truth is to establish a *correct answer* which can be used to test the SMoS. This known safety ranking needs to be established by previous studies using other methods. Once the ground truth has been established, a SMoS study can be conducted. The SMoS diagnosis aims to test whether a certain indicator (using different threshold values) can be used to observe the safety ranking established by the ground truth. To properly study the effect of a specific indicator (using a specific threshold), the SMoS study should use an event-based definition of exposure when counting the number of opportunities for a crash. If several indicators/thresholds produce the correct safety rankings, the most preferable option is the indicator/threshold with the highest observed frequency.

A case study looking at bicycle crossing design

The following section provides an example of a validation study of the TTC_{min} and PET using the proposed relative approach to validity. The study focuses on interactions between right turning motor vehicles and cyclists in signalized intersections with separated cycle crossings or cycle lanes (see Figure 16). Specifically, the study focuses on a subset of the intersections described in chapter 5 of this thesis. Three intersections with separated cycle crossings located in the Netherlands were chosen together with three intersections with cycle lanes located in Denmark. The Netherlands and Denmark were chosen because of their high number of cyclists and the consistent design of their respective intersections. Note that one of the locations in the Netherlands were not included because it used a slightly different design.

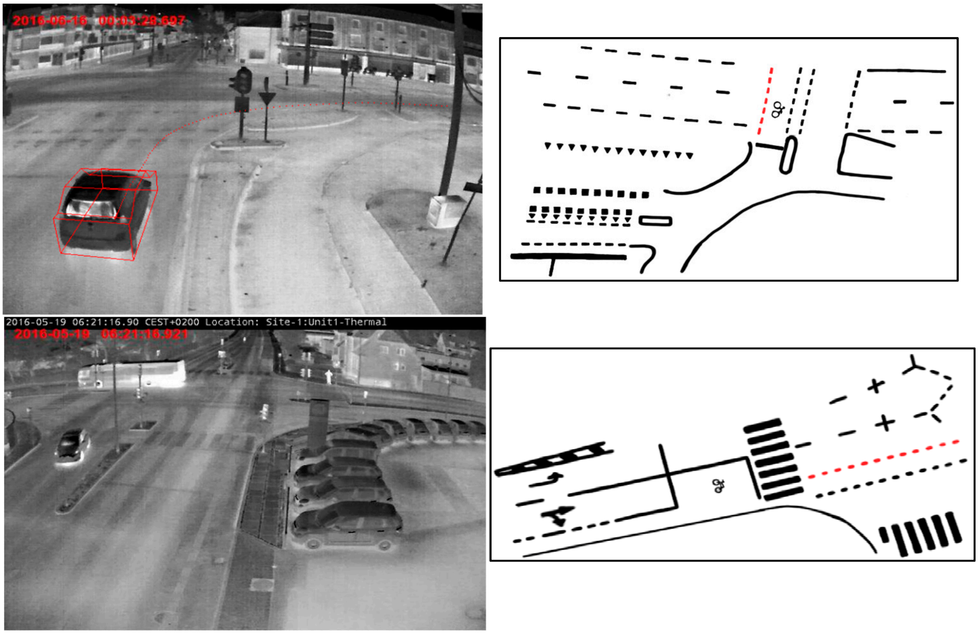


Figure 16. The two designs: separated cycle tracks (top image), and cycle lane (bottom image). The left images show the camera views at two intersections, and the right images show a schema of the design. The red dotted line shows were the speed measurements were made.

Safety ranking – ground truth

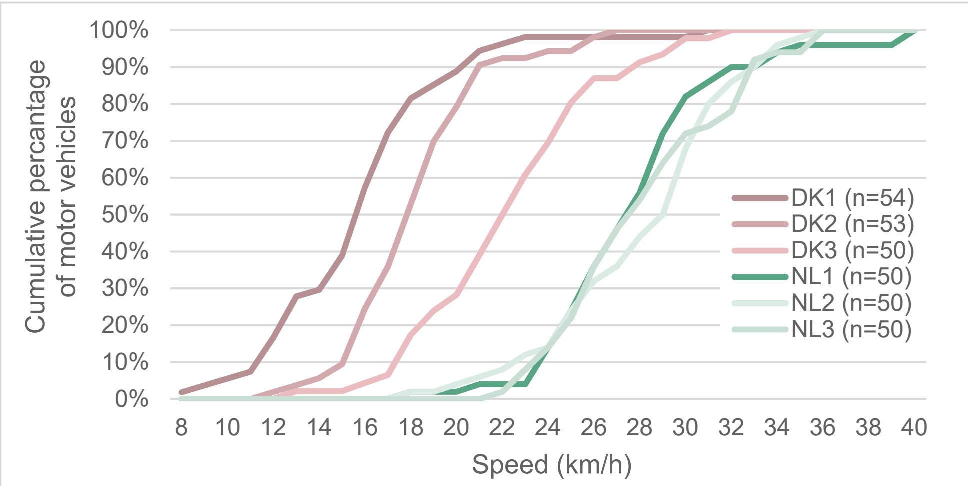
Separated cycle crossings are defined by a recessed cycle track in conjunction with a pedestrian crossing, in which the cyclists are physically separated from the motor vehicle before and after the intersection. In contrast, cycle lanes are only separated from the motor vehicles with painted lines which continue before and after the intersection.

According to the systematic literature review made by Prati et al. (2018), there is evidence that separated cycle crossings are less safe compared to cycle lanes. This result also seems to hold true when focusing solely on right turning motor vehicles (Jensen & Transportation Research, 2008; Summala et al., 1996). However, other researchers note that there is a general lack of high-quality evidence as to the effect of cycling infrastructure on cycling collisions (Mulvaney et al., 2015; Prati et al., 2018).

Speed seems to be an exception to the general lack of high quality evidence, and several reviews have consistently found that lowering the speed has a positive impact on safety (Aarts & van Schagen, 2006; Mulvaney et al., 2015; Prati et al., 2018; Thomas & DeRobertis, 2013). Following these results, Figure 17 below shows the speed of unhindered motor vehicles making a right-turn at the studied intersections, measured when the motor vehicles first start to cross the cycle path (shown as the red line in Figure 16). The speed distribution shows that the intersections in the Netherlands have a significantly higher crossing speed compared to the intersections in Denmark.

In conclusion, both the design and the speed measurements taken at the locations indicate that the 3 studied intersections in the Netherlands are less safe in comparison to the 3 Danish intersections.

Figure 17. The cumulative speeds of unhindered motor vehicles at the observed locations, measured when the motor vehicle starts to cross the cycle path.



The SMoS diagnosis

The SMoS data from the 6 intersections is a subset of the 24-hour data described in chapter 5 of this thesis. Table 11 shows the number of encounters identified and the corresponding number of critical events. Note that the column $TTC_{min} < \infty$ indicates the number of encounters in which a collision course was identified during the encounter, regardless of the TTC value. In general, more encounters occurred in the Dutch intersections, with the first intersection having considerably more events than the others.

Table 11. The number of encounters and safety critical events, using different threshold values, found during 24 hours at the 6 different intersections.

Site	**Encounters**	**TTCmin < ∞**	**TTCmin <= 4s**	**TTCmin <= 3s**	**TTCmin <= 2s**	**TTCmin <= 1.5s**	**PET <= 5s**	**PET <= 4s**	**PET <= 3s**	**PET <= 2s**	**PET <= 1s**
DK 1	205	28	23	18	7	3	195	193	189	157	54
DK 2	179	63	45	23	5	3	179	179	176	165	64
DK 3	205	17	10	7	3	2	195	195	195	179	54
DK *total*	589	108	78	48	15	8	569	567	560	501	172
NL 1	317	219	185	112	29	6	292	284	272	223	62
NL 2	190	90	84	58	20	11	188	185	177	137	27
NL 3	155	56	50	34	4	3	151	150	141	111	30
NL *total*	662	365	319	204	53	20	631	619	590	471	119

Taking the exposure levels at the different locations into consideration, Table 12 shows the number of critical events per encounter for the various threshold values. Table 12 also shows the p-value generated from Welch's t-test when comparing the Danish and Dutch results. Looking at the risk estimates in Table 12, the two indicators do not agree. Only using TTC_{min} does the SMoS analysis agree with the ground truth. Looking only at the mean value, this result for TTC_{min} is consistent for all threshold values. However, when looking at the individual sites, only the threshold values of 4s and 3s show that the intersections in the Netherlands are consistently less safe than those in Denmark. The same is found using the Welch's t-test, which generates p-values less than 0.05 for TTC_{min} with thresholds 3s, 4s and ∞.

Table 12. The number of critical events per encounter, using different threshold values, found during 24 hours at the 6 different intersections.

Site	TTCmin < ∞	TTCmin <= 4s	TTCmin <= 3s	TTCmin <= 2s	TTCmin <= 1.5s	PET <= 5s	PET <= 4s	PET <= 3s	PET <= 2s	PET <= 1s
DK 1	0.14	0.11	0.09	0.03	0.01	0.95	0.94	0.92	0.77	0.26
DK 2	0.35	0.25	0.13	0.03	0.02	1.00	1.00	0.98	0.92	0.36
DK 3	0.08	0.05	0.03	0.01	0.01	0.95	0.95	0.95	0.87	0.26
DK *Mean*	**0.19**	**0.14**	**0.08**	**0.03**	**0.01**	**0.97**	**0.96**	**0.95**	**0.85**	**0.29**
NL 1	0.69	0.58	0.35	0.09	0.02	0.92	0.90	0.86	0.70	0.20
NL 2	0.47	0.44	0.31	0.11	0.06	0.99	0.97	0.93	0.72	0.14
NL 3	0.36	0.32	0.22	0.03	0.02	0.97	0.97	0.91	0.72	0.19
NL *Mean*	**0.51**	**0.45**	**0.29**	**0.07**	**0.03**	**0.96**	**0.95**	**0.90**	**0.71**	**0.18**
P-value, Welch's t-test	**0.03**	**0.02**	**0.01**	**0.07**	**0.13**	**0.41**	**0.30**	**0.07**	**0.04**	**0.03**

The result from TTC_{min} is somewhat unexpected. SMoS theory would state that the stricter threshold values should provide a more accurate safety analysis. One potential explanation for this discrepancy is the significant decrease in the number of events which are selected using the threshold values of 2s and 1.5s. It is possible that the low number of events is creating a large random variation which makes the result somewhat unclear. It is also noteworthy that the result is consistent with the Scandinavian validation result discussed in the previous chapter of this thesis.

In contrast to TTC_{min}, the SMoS analysis using PET shows two major concerns. Firstly, there seems to be a strong correlation between encounters and critical events when applying threshold values larger than 2 seconds. This might indicate that for at least these high threshold values, the PET indicator is not really measuring safety, but mostly measuring exposure. The second concern is that the result from the SMoS analysis disagree with the ground truth. Using both a threshold value of 2s and 1s, the Danish sites are consistently found to be less safe than those in the Netherlands.

Discussion

This chapter argues that a relative approach to validity can allow for less resource intensive validity tests of SMoS. The proposed approach can be applied to comparatively quickly test a SMOS, without the need for any resource intensive work focusing on establishing the expected crash frequency. This is especially relevant in cases where the studied indicators fail to show the safety ranking expected by the ground truth. However, the proposed approach also has a number of limitations that are further discussed in this section.

The first limitation is that any validity test made based on previous studies rely on the validity and reliability of the studies used for the ground truth. Looking at the case study, the ground truth relies on several literature reviews that summarize existing knowledge about how safety relates to bicycle infrastructure design. In general, there seems to be an agreement about speed but some doubt regarding the infrastructure design itself.

The second limitation of the proposed approach is that there is no real yardstick for measuring how many locations are needed. It is obviously preferable for the SMoS study to include many locations; however, the suggested approach also works when studying fewer locations. This problem might be alleviated by an increased number of validation studies. Since the main advantage of the relative approach is that it allows for less resource intensive studies, it would hopefully result in an increased number of studies being performed. More, smaller, validity tests would also allow the indicators to be tested in differing scenarios, which is important if the aim of an indicator is to be universal.

The third and final limitation is that a relative approach is unable to analyse the absolute difference in frequency of critical events. For example, if a specific design produces twice the number of critical events per encounter, it would be impossible to infer that the risk is also twice as high. Instead, the only result would be that the design with a higher frequency of critical events is more dangerous. A response to this limitation could be to divide the concept of a SMoS study into two separate types of studies: (1) the classical SMoS study and the (2) relative SMoS study. The classical approach would focus on estimating the frequency of safety critical events with low threshold values, following the idea of the safety hierarchy (i.e. lower threshold values are *closer* to crashes), while the relative approach would focus on the highest possible threshold which still allows for the correct safety ranking analysis. This distinction makes theoretical sense in that the optimal threshold would be different for the two types, and practical sense in that it would allow for a shorter and more applicable relative approach as well as a longer, more resource intensive, classical approach to SMoS.

8. An exploration of encounters

One of the main findings so far is the strong correlation between encounters and SMoS. While this connection is not that surprising, since safety critical events by their nature will be a subset of all events, it does merit some further exploration of the encounters and other event-based exposure measurements.

The aim of this chapter is to further explore how encounters can (and should) be counted, and to explore the potential use of encounters themselves as a means of studying traffic safety. Note that these definitions are meant to provide a further exploration of the definition of *encounters*, and expand on the definition previously described in chapter 5.

The main part of this chapter focuses on the findings from paper 4. The aim of paper 4 is to explore whether encounters might help to explain the Safety in Numbers (SIN) effect in a similar fashion to the so called *Safety in Density* explanation (E. Thompson et al., 2015; J. Thompson et al., 2015; Thompson et al., 2016; Thompson et al., 2017; Thompson et al., 2018).

Three encounter definitions

As noted on page 16 in this thesis, an elementary unit of exposure is *any clearly defined and countable event that generates an opportunity for a crash to occur*. The idea is to only include the actual opportunities for crashes. However, since it is not obvious which traffic events create crash opportunities, this study uses three different definitions of encounters, as illustrated in Figure 18.

Type 1 relates to the number of simultaneous arrivals of conflicting road users at the conflict area. Whenever a motor vehicle and a cyclist are present within the intersection and their paths are expected to cross, it is considered an encounter. The situation in Figure 18 includes six Type 1 encounters, since both cars *encounter* each of the three cyclists.

The second type of encounter (Type 2) builds on the first, but considers the queue of motor vehicles and only counts the first motor vehicle in the queue. The rationale for excluding the other motor vehicles is that they are not really interacting with the cyclists and therefore do not pose a crash threat to the passing cyclists. The Type 2 situation visualised in Figure 18 includes three encounters, since the front-most car *encounters* each of the three crossing cyclists.

Figure 18. Visualisation of the three types of event-based exposure measures used in this study (Johnsson et al., 2020).

The third type (Type 3) of encounter continues to expand on the second type, only including one of the passing cyclists. The idea is that after a motor vehicle has interacted with, and possibly yielded for, a cyclist, the remaining cyclists that cross while the motor vehicle is motionless face no crash risk. Note that if the motor vehicle attempts to continue but must stop again for another cyclist, this results in a separate encounter. The obvious exception to this idea is that it is possible for several cyclists to be hit by a single motor vehicle. However, between 2014 and 2017, the Swedish police reported 2985 injury-causing crashes between motor vehicles and cyclists at intersections, and only 20 of these crashes involved more than one injured person. While this indicates that the risk for the cyclists who pass through the intersection after a motor vehicle ahead of them has already yielded is not zero, which also means that using an exposure measure that overlooks them might introduce bias, it is possible, keeping that in mind, that the Type 3 encounter can still provide some insight into the remaining 99.77% crashes.

There are two additional noteworthy aspects of the Type 3 encounter. First, it is not necessarily the first cyclist in *a group* who is *at risk*; instead, the point of this distinction is that only *one* of the cyclists in *the group* is at any considerable risk. Second, the number of passing cyclists is not the same as a *social group*, since it can include cyclists coming from both directions, and the division of where a new *group* begins is dependent on the behaviour of the motor vehicle. A new encounter initiates only if a previously stationary motor vehicle starts to move and then interacts with another cyclist.

Encounters and traffic volume

Manual observations were made during at least 24 hours at four of the intersections described in chapter 5 (two intersections in Sweden, one in Norway, and one in Denmark). To then analyse the relation between encounters and traffic counts, the data was divided into 15-minute periods and a Poisson regression model was fitted for both right- and left-turning motor vehicles to the observed data. The model form (Equation 2) selected is similar to the one commonly used to study the relationship between crashes and the volume of vulnerable road users (Elvik & Bjørnskau, 2017), with the addition of categorical predictor variables for the countries.

$$\frac{Encounters}{15\ minutes} = e^{\alpha} * \frac{Cyclists}{15\ minutes}^{\beta_{cycl}} * \frac{MVs}{15\ minutes}^{\beta_{mv}} * e^{\alpha_{den}*Den+\alpha_{nor}*Nor+\alpha_{swe}*Swe} \quad (2)$$

Table 13. Coefficients from the Poisson regression model for right-turning MVs (95% Wald CI). Taken from Johnsson et al. (2020).

Parameter	Right-turning motor vehicles		
	Type 1	Type 2	Type 3
$\boldsymbol{\alpha}$ *(intercept)*	-4.18 (-4.51 to -3.853)	-3.59 (-4.01 to -3.17)	-2.98 (-3.47 to -2.48)
$\boldsymbol{\beta_{cycl}}$	**1.30 (1.24 to 1.35)**	**1.08 (0.99 to 1.15)**	**0.64 (0.54 to 0.74)**
$\boldsymbol{\beta_{mv}}$	**0.98 (0.88 to 1.07)**	**0.79 (0.67 to 0.90)**	**0.81 (0.66 to 0.95)**
$\boldsymbol{\alpha_{den}}$	-0.88 (-1.05 to -0.72)	-0.41 (-0.61 to -0.21)	0.18 (-0.07 to 0.43) *
$\boldsymbol{\alpha_{nor}}$	-0.67 (-0.87 to -0.47)	-0.23 (-0.46 to -0.01)	0.00 (0.26 to -0.26) *
$\boldsymbol{\alpha_{swe}}$	0 (reference)	0 (reference)	0 (reference)

Table 14. Coefficients from the Poisson regression model for left-turning MVs (95% Wald CI). Taken from Johnsson et al. (2020).

Parameter	Left-turning motor vehicles		
	Type 1	Type 2	Type 3
$\boldsymbol{\alpha}$ *(intercept)*	-4.60 (-4.88 to -4.32)	-3.48 (-3.80 to -3.16)	-2.57 (-2.96 to -2.18)
$\boldsymbol{\beta_{cycl}}$	**1.33 (1.25 to 1.41)**	**1.09 (0.99 to 1.19)**	**0.57 (0.45 to 0.69)**
$\boldsymbol{\beta_{mv}}$	**1.04 (0.96 to 1.12)**	**0.71 (0.62 to 0.80)**	**0.69 (0.57 to 0.81)**
$\boldsymbol{\alpha_{den}}$	-0.49 (-0.57 to -0.40)	-0.14 (-0.24 to -0.03)	0.22 (0.07 to 0.37)
$\boldsymbol{\alpha_{nor}}$	-1.40 (-1.91 to -0.90)	-1.47 (-1.99 to -0.94)	-1.08 (-1.68 to -0.49)
$\boldsymbol{\alpha_{swe}}$	0 (reference)	0 (reference)	0 (reference)

The resulting parameters (Table 13 and 14) show that for Type 1 encounters, the number of encounters seems to increase faster than the increase in volume (i.e. the coefficients are larger than 1), or at least remain proportional to the traffic volume. This result fits with the predictions made by Rune Elvik et al. (2009).

However, the results for the Type 2 encounters show a less than linear relation to motor vehicle volume but not to cyclist volume, and Type 3 encounters clearly show a less than linear relationship between both types of volume and encounters, meaning that an increase in volume does not correspond to a proportional increase in the number of encounters. The results for the Type 3 encounters especially, produce a clear SIN effect between volume and encounters, like the one commonly found between crashes and volume. This result from both Type 2 and Type 3 encounters can be explained by an increase in the mean queue length and an increase in the average size of the cyclist *groups*, which occurs as a consequence of a higher traffic flow which creates a non-linear relation between traffic flow and encounters.

Encounters and crashes

Based on the result from the previous section, a further attempt to correlate encounters and crashes was made using Type 3 encounters. Looking back at the crash models developed in chapter 6 using the Scandinavian crash data (Table 7), the cyclist model showed a typical SIN effect. If a linear correlation could be found between encounters and crashes, this would further strengthen the idea that they could help to explain the phenomenon.

To be able to compare the crash model based on volume to the crash model based on encounters, an estimation of the average daily number of encounters had to be made for the 166 different locations for which crash data was collected. The estimation was made based on the traffic counts for each location and the established relationship between encounters and traffic volume presented in the previous section. Since the aim was to test whether the relationship between crashes and encounters is linear, a normal crash model was then developed following the methodology used for the previously established crash models using the following model form:

$$\frac{Crashes}{year} = e^{a} * \frac{ENC^{b}}{day} \ , \qquad (3)$$

where $Crashes$ is the expected number of crashes per year and ENC is the estimated daily number of encounters. Table 15 below shows the resulting regression parameters. The estimate of 0.9 is quite close to the expected value of 1 where the relationship between encounters and crashes is linear. However, as in the previous model, the standard error is quite large, and it is difficult to draw any robust conclusions.

Table 15. Value of regression parameters for the crash model based on encounters. Taken from Johnsson et al. (2020).

Parameter	Estimate	Standard error	Wald 95% confidence limits	Pr > ChiSq
α *(intercept)*	-6.74	1.35	-9.40 – -4.09	< .0001
β (enc)	0.90	0.23	0.45 – 1.36	< .0001
Dispersion	0.81	0.82	0.11 – 5.59	-

Discussion

From the point of view of encounters as an explanation for the SIN effect, the results from the study indicate that the relationship between traffic volume and encounters shows a SIN effect similar to that which is normally found between volume and crashes, when applying the Type 3 encounter, i.e. encountering *groups of road users*. The crash model also indicates that it is not infeasible that a linear relationship between crashes and encounters might exist. However, the limited number of locations and the limited crash data makes any strong conclusions impossible; further research with more data is needed.

From the point of the research questions explored in this thesis, the main result from the paper is that event-based exposure alone might have some explanatory power when it comes to safety studies. If this hypothesis of protected road users is valid, observing and analysing this effect could itself be of value. This might present an opportunity for studies focusing solely on event-based exposure as a method for analysing at least some parts of traffic safety.

There are also some noteworthy limitations with the study and its use of encounters. While the crash data from the Swedish accident database lends some credence to the argument that certain cyclists pass *unexposed* because of the actions of a prior cyclist, it is still unclear exactly how the risk differs between different cyclists who pass in front of a motor vehicle. Further studies into how the risk of cyclists is affected by when they pass in front of a motor vehicle might be able to better understand this process.

Another aspect that has not been investigated in this study is the complexity and ease of use of encounters. Compared to traffic volume, encounters are considerably more complex to identify and count, which might limit their practical usefulness. Whether this increase in complexity is worth the improved insight into crash causality is an open question. One simple step to attempt to answer this is to estimate the number of encounters based on observed traffic volume. However, this method relies on making robust estimations between encounters and traffic volume in the first place. Future research could focus on establishing such estimations for different types of infrastructure, which could then be used in conjunction with traditional traffic counts instead of having to directly identify and count the encounters. Since the encounters are defined solely by spatial rules, micro-simulation could be used to study this relationship in a wide range of different scenarios.

9. Detecting the beginning of evasive actions

This chapter will present an algorithm for detecting the start of an evasive action, based solely on trajectory data, followed by a discussion on how the same approach can be extended to predict how a road user would have continued to travel if no evasive action was taken.

The motivation for this work is twofold. Firstly, the result and discussion from the literature review part of this thesis show a potential benefit to estimating severity before the start of an evasive action instead of after the event has taken place; however, to my knowledge, no previous research has attempted to identify the start of an evasive action using solely trajectory data. The studies discussed in the literature review which use either the Swedish conflict technique or only the Time to Accident (TA) indicator all rely on a human observer to identify the start of an evasive action.

The second motivation is the need for a more robust motion prediction approach. There are several types of motion prediction techniques which could be used to find the future path of road users. The survey by Lefèvre et al. (2014) divides the motion prediction techniques into three categories:

1. Physics-based motion models are the simplest models, they consider that the motion of vehicles only depends on the laws of physics.

2. Manoeuvre-based motion models are more advanced, as they consider that the future motion of a road user also depends on the manoeuvre that the road user intends to perform. These models can involve trajectory learning, in which empirical observations are used to form the basis for the future motion.

3. Interaction-aware motion models are the most advanced, and consider the inter-dependencies between the road users' manoeuvres.

As previously discussed in the literature part of this thesis, the most common approach to motion prediction in SMoS studies is simple physics-based motion models (Laureshyn et al., 2016; Aliaksei Laureshyn et al., 2010; St-Aubin et al., 2014). There are also some attempts at using motion patterns in a manoeuvre-based motion model to calculate TTC (St-Aubin et al., 2014). These approaches attempt to predict how the road user will continue travel if the road user remains on/at its current path and speed. However, these models risk producing unrealistic predictions if they are used to calculate how a road user will act once it has started to interact with other road users, since neither physics-based nor manoeuvre-based motion models consider any interacting behaviour.

Methodology

This section will present a general method for identifying the start of an evasive action from any road user based on trajectory data, and how a simple manoeuvre-based motion model can then be used to make motion predictions from the moment before an evasive action is detected.

The basic idea of the method relies on separating *unhindered* from *interacting* trajectories by using a comparison set of trajectories from unhindered road users, and calculating how similar a specific trajectory is to that set. If a trajectory is significantly different from the unhindered set at any point in time, this indicates that the considered trajectory has stopped being *unhindered* and is therefore *interacting* with another road user.

To calculate the similarity between trajectories at a specific moment in time, this study proposes a simple method that relies on the average distance between two trajectories. The calculation of *similarity* at a timestep is made in two steps. Firstly, the closest point between the current position and the entire unhindered trajectory is identified. Secondly, using the closest point as a starting point, the distances between the points ($\Delta s\ in$ Figure 19) in the two trajectories can be calculated. The final *similarity* at timestep t is the average distance calculated from the current position/closest position and n timesteps backwards in time as shown in the equation below.

$$\text{Similarity}_t = \frac{\sum_1^n \Delta s}{n} \qquad (4)$$

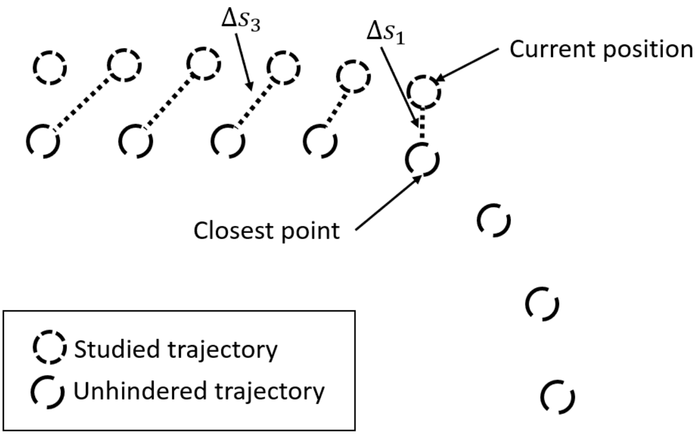


Figure 19. Based on the current position in the trajectory and the corresponding closet position in the unhindered trajectory, the average distance (i.e. similarity) is calculated looking n timesteps back.

By using a set of unhindered trajectories and calculating the number of similar trajectories for each time-step, it becomes possible to determine at what time-step an interaction starts, i.e. when there are no more similar trajectories.

Following the identification of an evasive action, the already established similarity concept can be used in a manoeuvre-based motion model by assuming that the studied road users would have continued to travel in the same way as the trajectories considered to be similar. This approach allows for several potential future paths depending on the number of similar trajectories at each time-step. However, this prediction is only possible while there are similar trajectories, meaning that the prediction cannot be made once an evasive action has been identified. The prediction must therefore be made right before the start of an evasive action.

Using the motion model, it is possible to estimate the probability of the road users being on a collision course. It is also possible to calculate how far into the future any collisions would occur, i.e. the time to collision right before the starting point of evasive action.

Experimental data

The data used for the following tests come from the 1-day data gathered from 7 intersections (Figure 20) described in chapter 5. Note that this experiment is limited to only include interactions between right-turning motor vehicles and cyclists. In addition to this data, a set of unhindered trajectories were gathered at each intersection.

Figure 20. The camera view at the seven intersections. Note that a thermal camera was used in Denmark and in the Netherlands.

Calibration

The proposed algorithm for calculating *similarity* has two parameters that influence the result: 1) the threshold for the average distance between the trajectories, and 2) a limit on how far into the past the calculation should be made.

To find suitable parameter values, a calibration test was conducted using traffic events which were manually selected from the dataset. A total of 50 interactions between right-turning motor vehicles and cyclists were selected. Each of the selected situations involved a distinct and clear evasive action. Following the identification of these events, four other researchers experienced in watching and evaluating traffic situations were asked to identify the start of the first evasive action.

To test how well the algorithm agreed with the human observers, the Intraclass Correlation Coefficient (ICC) was used to measure the reliability among the different observers (Fisher, 1932). The ICC produces a reliability index between 0 and 1 when comparing the result from different raters. Values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively (Koo & Li, 2016). There are many different forms of the ICC index. Following the guideline by Koo and Li (2016), a Two-Way Mixed-Effects Model focused on absolute agreement was chosen for the test.

By testing various combinations of possible parameter values, the best result was found using an average distance value of 0.8m and a time parameter of 2.93s. With these parameters, the ICC values showed a good to excellent reliability regardless of which test person was compared to the computer result. However, the reliability between the observers themselves was slightly better than when compared to the computer result. In addition to the ICC values, it is also interesting to analyse how accurate the computer was when compared to the mean result from the observers. In this case, the algorithm was generally slightly early in its detections, with a mean error of -0.16s. The computer also had a noticeably higher standard deviation of 0.73s when compared to the human observers which had a standard deviation between 0.25-0.32s. However, the result overall indicates that the algorithm does a satisfactory job at identifying the first evasive action with good to excellent reliability when compared with human observers.

A test at seven intersections

The following section presents the result of analysing the 1099 encounters between right-turning motor vehicles and cyclists observed at the 7 intersections. Following the structure of the proposed method, each interaction can be classified into four main categories:

1. Events with no detected evasive action
2. Events with a detected evasive action without a probability of collision course (PCC)
3. Events with a detected evasive action and a non-zero PCC
4. Abnormal and secondary events

The first three categories follow from the previous method description; however, the fourth type of events were identified when analysing the result. These abnormal events are defined as being immediately detected as evasive actions the moment both road users are visible in the camera view. In these cases, the algorithm is unable to make any motion predictions, since no similar trajectories were ever detected. The result suggests two different types of situations that can lead to these events. The first type of abnormal event is situations in which one or both road users showcase uncommon behaviour that is too different from the behaviour of the unhindered trajectories that are used by the algorithm. The second type of abnormal event is secondary interactions, in which one of the road users has already interacted with another road user before the second road user has entered the camera view. The algorithm correctly identifies that an evasive action has occurred the moment in which the second road user enters the camera view, but cannot make any motion predictions from that point.

Table 16 shows the result separated into the four categories. There are several noteworthy results. Firstly, it is quite uncommon that an encounter occurs without any form of evasive action; however, it is significantly more common at the Danish sites. Secondly, only 14% of all events have a non-zero chance of being on a collision course and therefore produce a TTC value, compared to the 35% of encounters with a TTC value using the classical approach to TTC (see the discussion starting on page 60). Finally, the algorithm fails in 19% of all events, either due to abnormal behaviour or due to secondary interactions. The frequency of failed measurements also seems to have a high variation with a maximum of 53% and a minimum of 9%.

Table 16. The result divided into four separate categories.

Intersection	Encounters	Category 1	Category 2	Category 3	Category 4
DK 1	80	29 (37%)	26 (33%)	4 (5%)	19 (24%)
DK 2	137	21 (15%)	85 (63%)	18 (13%)	12 (9%)
DK 3	114	21 (19%)	49 (44%)	5 (4%)	37 (33%)
NL 1	107	4 (4%)	49 (46%)	20 (19%)	34 (32%)
NL 2	142	1 (1%)	97 (68%)	30 (21%)	14 (10%)
NL 3	109	0 (0%)	39 (36%)	12 (11%)	57 (53%)
Spain	416	18 (4%)	297 (71%)	60 (14%)	41 (10%)
Total	**1099**	**94 (9%)**	**642 (58%)**	**149 (14%)**	**214 (19%)**

Table 17 and Table 18 show more detailed data from the encounters which produced a non-zero PCC and therefore also a corresponding TA value (i.e. Category 3). Note that the TA values in Table 18 correspond to the mean TA value from all combinations of the motion prediction that resulted in a collision between the trajectories. There are two main findings from these tables. Firstly, the PCC values seem to be quite constant, with the only exceptions being sites Denmark 1 and 3. These sites are also noteworthy in that they produce significantly fewer values compared to the remaining sites. One potential explanation for this is the camera angle, which captures significantly less of the incoming path towards the intersection at both these sites (see Figure 20). The second finding is that the Danish sites seem to generally produce lower TA values compared to the other locations. However, there also seems to be fewer events at the Danish sites.

Table 17. Summary statistics of the probability of collision course from the 7 intersections.

PCC	DK 1	DK 2	DK 3	NL 1	NL 2	NL 3	Spain
Events	4	18	5	20	30	12	60
Mean	0,22	0,38	0,21	0,32	0,32	0,33	0,34
St. dev.	0,15	0,31	0,16	0,24	0,22	0,21	0,26
Min	0,03	0,04	0,07	0,02	0,03	0,06	0,02
Max	0,39	1,00	0,46	1,00	0,83	0,80	1,00

Table 18. Summary statistics of the mean Time to Accident indicator from the 7 intersections.

TA	DK 1	DK 2	DK 3	NL 1	NL 2	NL 3	Spain
Events	4	18	5	20	30	12	60
Mean	1,64	1,73	1,32	2,52	2,31	2,43	2,42
St. dev.	0,58	0,71	0,17	0,69	0,54	0,60	0,97
Min	1,20	0,94	1,13	1,40	1,32	1,63	0,05
Max	2,49	3,80	1,52	3,68	3,37	3,45	5,13

Discussion

The result shows that the algorithm demonstrates potential in its ability to detect the start of an evasive action. It clearly shows a distinct difference between the safety critical events and the normal meetings. However, the result also shows that the algorithm cannot properly handle secondary interactions. Overall, the proposed method for identifying evasive actions, and the motion model, can be suitable for use in practice. However, further research into how to properly identify and analyse secondary interactions is needed, as well as a framework for how to handle situations without evasive actions and how these events should be merged into a single safety analysis.

The calibration test indicated that the best parameter values were an average distance of 0.8m looking at least 2.9s into the past. However, there are some potential concerns with this result. First, it is possible that the suitable values are dependent on both the camera view and the technical processing of the video. For example, the angle and height of the camera at the first and third site in Denmark are the likely causes of the lower number of events with calculable values there. It is also possible that changing what type of camera calibration is used and the accuracy of the tracking software will affect suitable parameter values (T-Analyst uses the Tsai camera calibration method (Tsai, 1986), and the tracks are manually made by the user).

Considering the previous chapter discussing relative validity, it is noteworthy that the result shows considerably lower (but fewer) TA values at the Danish sites compared to the Dutch intersections. However, it is difficult to draw any concluding remarks due to the problem with camera angles discussed in the previous section. If anything, the result might indicate that using TA alone might not be suitable as a SMoS, and that adding other consideration such as speed or deceleration might produce different results.

Finally, from the point of the research questions explored in this thesis, it is interesting to note the aspect of secondary interactions, which limits the use of the proposed algorithm. The question of how frequently critical events are also secondary interactions is interesting. It makes some intuitive sense that secondary interactions are more dangerous in comparison to an average event, since such situations involve multiple moving road users, which increase the complexity and therefore also the risk. However, it is also possible that it is the first interaction that is the most critical, and any secondary interaction is mostly safe due to the already heightened attention of the road users and the generally lower speeds after the first interaction.

10. Final discussion

SMoS, VRUs and validity

Previous validation studies have generally shown a strong relation between crashes and critical events. However, most of these studies focus on MVs and rely on human observers. The attempt at classical validation made in this thesis also shows a significant relation between crashes and critical events for both TTC_{min} and PET, but the attempt also showcases several difficulties of such an enterprise. The main problem is estimating the crash frequency, which can be compared to the frequency of critical events. This is not only highly resource intensive, but the problem of underreporting and the very low number of crashes (when separated into manoeuvres) make this a very difficult task.

There are some potential workarounds for this problem; it might be easier to focus solely on MVs, to limit the data collection to only one country, and to include the entire intersection and not only one leg of an intersection. These changes might allow for an easier estimation of crash frequency; however, it is unclear whether the result would be viable for uses in other countries or with VRUs.

A further problem is the lack of an answer to the question: *how strong must the relation be between crashes and critical events for them to be considered valid?* This problem illustrates the trade-off between shorter observation periods and a stronger relation to traffic safety. Assuming a SMoS correctly measures severity, a more severe threshold value should result in a more robust analysis of traffic safety but will also result in fewer observed critical events, which in turn requires longer observation periods to produce robust results.

The relative approach to validity and the separation of SMoS studies into short-term and long-term can solve some, but not all, of these problems. The relative approach to validity focuses only on whether the SMoS can identify the relative safety difference between two traffic scenarios instead of the difference in the frequency of critical events. This allows for a less resource intensive type of validity study with no estimation of crash frequency and less studied locations overall. This approach also provides a clear yardstick which can be used to find both best SMoS and the optimal threshold value for each SMoS. The best indicator will be the one that is able to correctly identify the relative difference in safety while maintaining the highest critical event frequency, and therefore the shortest possible observation period. This SMoS and its threshold value can then be used in short-term SMoS studies which focus on whether a safety improvement has been observed between two options.

However, the relative approach has an obvious weakness in that it cannot measure how much of a difference in safety has been observed, or more specifically, it cannot provide any support that the observed difference has any relation to the expected change in crash frequency. It is therefore also important that a long-term type of SMoS study be used to investigate this question. Note that since we expect the relation between critical events and crashes to strengthen with a more severe threshold, the best possible threshold for estimating crash frequency must, in this context, be the most severe threshold value.

The EVT (Extreme Value Theory) approach to SMoS studies might provide an attractive prospect for long-term SMoS studies. The EVT approach attempts to directly estimate the crash frequency from critical event observations by calculating the frequency of situations with a threshold value that in practise indicates that a crash has occurred (for example, a crash occurs if the TTC value is 0 or less). However, while this approach can directly estimate the crash frequency, it has two main downsides. The first downside is the need for many observations of very severe situations (Lägnert, 2019), which in turns leads to longer observation periods. The second downside is that it does not provide a clear yardstick for which indicator provides the best result. For example, a value of 0 indicates a crash for both PET and TTC, however, the result from chapter 4 shows that lower PET values are considerably more common compared to TTC. Using the EVT approach will likely produce different crash frequency estimates depending on which indicator is used, and it is unclear how to identify the *best* one. Indeed, it is possible that this leads to a similar trade-off problem as the one described before. Using TTC might likely produce a better estimate compared to PET which in turn might need a considerably shorter observation period. This discussion leads to the original question of how strong of a relation there must be before the indicator is considered valid.

SMoS and VRUs in practise

The issue of observation period and using SMoS in practise has to some extent been discussed in the previous section. However, the concern of false positives and the need for human observers in the loop are also of major importance.

The result suggests that the indicators used in the InDeV project produce too many false positives to be directly used without further consideration. The original idea that critical events identified by humans could be used to estimate the frequency of severe events failed when a larger number of severe events were found in the datasets containing normal events. This result implies either that the indicators (regardless of threshold value) fail to properly capture severity, and incorrectly identify some situations as severe, or that the human observers failed to include many severe situations.

Assuming the indicators fail and produce too many false positives, there are two possible solutions: a human selection somewhere in the study process, or an indicator which does not produce as many false positives. This could either mean a human pre-selection, as mostly used in InDeV, or an after-selection, in which an automated tool is used to find potential critical events, and a human then removes the false positives. Note that a human observer in the loop might also make the result more comparable to the various validation studies from the literature.

It is probably not necessary to create a perfect indicator without any false positives. However, since severe situations are quite rare, and there is a very large number of normal situations, even a small share of false positives risks overwhelming the low number of severe events. There is also a risk that a SMoS becomes a surrogate to exposure and not safety if the indicator produces many false positives. This effect might to some extent hide the fact that the SMoS is not working by instead relying on the useful characteristics of event-based exposure.

Motion prediction

Several SMoS indicators rely on motion prediction. The prediction method mainly used in this thesis is based on an assumption of constant speed. Assuming that the road user will continue without changing their speed is a naïve approach that does not consider any changes brought by the infrastructure, nor by interactions with other road users. More advanced methods for future motion prediction (see for example Mohamed and Saunier (2013)) might alleviate some of these concerns. Their approach uses past behaviour at the location to predict how the road users might continue. This allows the prediction to consider how a road user usually travels, and make a prediction based on that. This approach should make better predictions from further away. However, once the road users have started to interact and/or are taking an evasive action, assuming that they will continue like *normal traffic* might lead to bad predictions.

Following this argument, a solution might be to rely on naïve and simple models for short term predictions (up to maybe 1 second), and to rely on more advanced motion prediction models for longer predictions. One option could be to first identify when a road user starts to interact, and at that point switch from a long-term to a short-term prediction.

Observation period

As discussed in chapter 7, there are some reasons for dividing SMoS studies into a long-term and short-term form. The long-term study would focus on directly estimating the crash-frequency at the location, with a focus on only the most severe events. The expected observation period for such a study would likely be several weeks.

The short-term SMoS study would be more similar to the way SMoS studies have been used in the literature, with observations from one or a few days. However, this approach would not claim to be able to estimate the crash-frequency, and instead only claim to tell whether the safety has improved/worsened in comparison to a similar study at a different location. These studies would have to rely on observing less severe events compared to the long-term approach. Since this could imply a larger risk of false positives and a stronger connection to event-based exposure, further validity studies are recommended in order to identify a suitable indicator and their respective thresholds.

The relative approach to validity discussed in this thesis should allow for much less resource intensive studies, which hopefully makes future validity focused research more viable.

SMoS, VRUs and exposure

The results in this thesis indicate that event-based exposure should be used in conjunction with SMoS instead of traditional traffic counts. This allows the SMoS to better analyse traffic safety, without the risk of the result being influenced by the inherent connection between critical events and encounters. Event-based exposure also provides a clear distinction between what is exposure and what is a SMoS. The event-based exposure attempts to identify the frequency of events which have a non-zero probability of a crash, while the SMoS attempts to estimate how large that probability is, depending on the severity of the event. By then estimating the frequency of severe events in relation to the number of total events, a clear description of the safety can be made.

However, as discussed in the exploration of encounters (chapter 8), more research is needed into how event-based exposure functions and is measured. Expanding on the idea of protected road users might be of particular interest for future research. Looking at the concept of *group* used in the thesis, the argument is that several passing cyclists are *protecting* each other, since it is highly unlikely that more than one of them would be involved in a potential collision, though there is nothing that says that such protection could not involve other road users traversing different parts of an area. Taking the situation showed in Figure 21 as an example, the straight going MV, bicyclists, and pedestrian are protecting each other from the left turning MV, i.e. if this situation were to result in a crash, it is exceedingly unlikely that the crash would involve more than one of them.

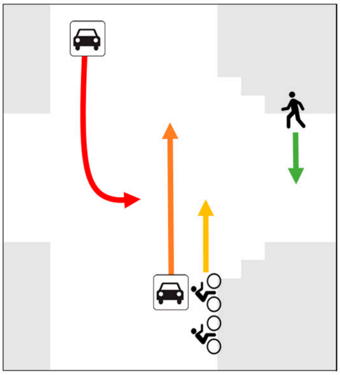


Figure 21. An example of protection.

This type of protection would lead to a much more complex definition of encounter which would depend on more than just the two interacting traffic flows. Furthermore, the situation in Figure 21 includes many potentially interacting road users, and this complexity might increase the risk of a collision occurring due to the many different road users that the left turning MV must pay attention to. This would mean that higher traffic complexity can both limit the number of opportunities for crashes, and also increase the risk for the opportunities that do exist. If so, it could provide a path for future analyses of traffic scenarios based on the aspect of complexity and protection. Since the analysis of encounters revolves around analysing normal conditions, an analysis of these aspects might be done quickly, while still providing relevant information related to the safety of the studied locations.

11. Conclusions

- There is a general correlation between SMoS and crashes in most previous validation studies. However, most of them focused on motor vehicles and relied on indicator estimates from human observers. The most used and best validated SMoS indicators are Time to Collision Minimum and Post-encroachment Time.

- Time to Collision Minimum seems to generally outperform Post-encroachment Time based on the limited classical validation study and the results from the relative validity study presented in this thesis. Furthermore, a threshold of 3-4 seconds for Time to Collision Minimum produced the best result in both studies.

- Both Time to Collision Minimum and Post-encroachment Time seem to produce a considerable number of false positives when applied to normal events in traffic and compared to human observers. Using an automated system to identify severe events followed by a human removing the false positives is the preferred solution until a more robust indicator is developed.

- Classical validation studies which focus on the correlation between crash- and critical-event frequency are very resource intensive. The problem of underreporting and the low number of crashes including VRUs further complicates attempts at validation. Relative validity studies which focus solely on establishing the relative difference in traffic safety between locations might be an attractive alternative. This approach is considerably less resource intensive and does not require an estimated crash-frequency.

- Event-based exposure should be used in conjunction with SMoS studies due to their inherent connection to critical events. However, how to identify and study event-based exposure should be further explored in future research.

12. References

Aarts, L., & van Schagen, I. (2006). Driving speed and the risk of road crashes: a review. *Accid Anal Prev*, *38*(2), 215-224. https://doi.org/10.1016/j.aap.2005.07.004

Alhajyaseen, W. K. M. (2015). The Integration of Conflict Probability and Severity for the Safety Assessment of Intersections. *Arabian Journal for Science and Engineering*, *40*(2), 421-430. https://doi.org/http://dx.doi.org/10.1007/s13369-014-1553-1

Allen, Shin, & Cooper. (1978). Analysis of traffic conflicts and collisions. *Transportation Research Record*, *667*, 67-74.

Alsop, J., & Langley, J. (2001). Under-Reporting of Motor Vehicle Traffic Crash Victims in New Zealand. *Accident Analysis & Prevention*, *33*(3), 353-359. https://doi.org/http://dx.doi.org/10.1016/S0001-4575(00)00049-X

Amoros, E., Martin, J.-L., & Laumon, B. (2006). Under-Reporting of Road Crash Casualties in France. *Accident Analysis & Prevention*, *38*(4), 627-635. https://doi.org/http://dx.doi.org/10.1016/j.aap.2005.11.006

Amundsen, F. H., & Hyden, C. (1977). Proceedings: First workshop on traffic conflicts. First workshop on traffic conflicts, Oslo, Norway.

Andersen, C., Kamaluddin, A., Várhelyi, A., Madsen, T., & Meltofte, K. (2016). *Review of current study methods for VRU safety. Appendix 7 –Systematic literature review: Self-reported accidents* (Deliverable 2.1 – part 4).

Asmussen, E. (Ed.). (1984). *Proceedings of the NATO advanced research workshop on international calibration study of traffic conflict techniques. 25-27 May 1983, Copenhagen, Denmark*. NATO ASI Series F: Computer and System Science vol. 5.

Bagdadi, O. (2013). Estimation of the severity of safety critical events. *Accident Analysis & Prevention*, *50*, 167-174. https://doi.org/http://dx.doi.org/10.1016/j.aap.2012.04.007

Baguley, C. J. (1984, 25-27 May 1983). *The British traffic conflicts technique* NATO advanced research workshop on international calibration study of traffic conflict techniques, Copenhagen, Denmark.

Baker, W. T. (1972). *An evaluation of the traffic conflicts technique* 51st Annual Meeting of the Highway Research Board,

Björnberg, A. (2016). *Do we learn more from situation based accident coding? Case study of Malmö and Gothenburg* (Publication Number Thesis 295) [Degree project, Lund University, Sweden].

Brinberg, D., & McGrath, J. E. (1985). *Validity and the research process*. SAGE Publications, Inc.

Brown. (1994). Traffic conflicts for road user safety studies. *Canadian Journal of Civil Engineering*.

Brown, Chau, A., DeCastilho, B., & Navin, F. (1984). *Traffic conflict procedure and validation*.

CaDaS. (2015). *Common accident data set. Referece guide* (Version 3.4 - April 2015).

Cafiso, S., García, A. G., Cavarra, R., & Rojas, M. A. R. (2011, 14-16 September 2011). *Crosswalk Safety Evaluation Using a Pedestrian Risk Index as Traffic Conflict Measure* 3rd International Conference on Road Safety and Simulation, Indianapolis, USA.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. SAGE Publications, Inc.

Chapman, R. (1973). The concept of exposure. *Accident Analysis & Prevention*, *5*(2), 95-110. https://doi.org/http://dx.doi.org/10.1016/0001-4575(73)90018-3

Chin, H. C., Quek, S. T., & Cheu, R. L. (1992). Quantitative examination of traffic conflicts. *Transportation Research Record*, *1376*, 86-91.

Davis, G. A., Hourdos, J., Xiong, H., & Chatterjee, I. (2011). Outline for a causal model of traffic conflicts and crashes. *Accident Analysis & Prevention*, *43*(6), 1907-1919. https://doi.org/http://dx.doi.org/10.1016/j.aap.2011.05.001

El-Basyouny, K., & Sayed, T. (2013). Safety performance functions using traffic conflicts. *Safety Science*, *51*(1), 160-164. https://doi.org/http://dx.doi.org/10.1016/j.ssci.2012.04.015

Elvik, R. (2015). Some implications of an event-based definition of exposure to the risk of road accident. *Accident Analysis & Prevention*, *76*, 15-24. https://doi.org/10.1016/j.aap.2014.12.011

Elvik, R., & Bjørnskau, T. (2017). Safety-in-numbers: A systematic review and meta-analysis of evidence. *Safety Science*, *92*, 274-282. https://doi.org/http://dx.doi.org/10.1016/j.ssci.2015.07.017

Elvik, R., Erke, A., & Christensen, P. (2009). Elementary Units of Exposure. *Transportation Research Record: Journal of the Transportation Research Board*, *2103*, 25-31. https://doi.org/10.3141/2103-04

Elvik, R., Høye, A., Vaa, T., & Sørensen, M. (2009). *The handbook of road safety measures*. Emerald Group Publishing Limited, Bingly, UK.

Elvik, R., & Mysen, A. (1999). Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. *Transportation Research Record*, *1665*, 133-140. https://doi.org/https://doi.org/10.3141%2F1665-18

Erke, H. (1984). *The traffic conflict technique of the Federal Republic of Germany* NATO advanced research workshop on international calibration study of traffic conflict techniques, Copenhagen, Denmark.

Evans, L. (2004). *Traffic Safety*. Science Serving Society.

Fisher, R. A. (1932). *Statistical methods for research workers*. Oliver and Boyd.

Grayson, G. B. (1984). *The Malmö study: a calibration of traffic conflict techniques* (R-84-12).

Güttinger, V. A. (1982, April 1982). *From Accidents to Conflicts: Alternative Safety Measurement* Third International Workshop on Traffic Conflicts Techniques, Leidschendam, The Netherlands.

Hansson, A. (1975). *Studies in driver behaviour, with applications in traffic design and planning. Two examples* (Bulletin 11).

Hauer, E. (1978). *Traffic conflict surveys: some study design considerations* (Supplementary report 352).

Hauer, E. (1982). Traffic conflicts and exposure. *Accident Analysis & Prevention*, *14*(5), 359-364. https://doi.org/http://dx.doi.org/10.1016/0001-4575(82)90014-8

Hauer, E. (1997a). *Observational before-after studies in road safety*. Pergamon.

Hauer, E. (1997b). *OBSERVATIONAL BEFORE-AFTER STUDIES IN ROAD SAFETY -- ESTIMATING THE EFFECT OF HIGHWAY AND TRAFFIC ENGINEERING MEASURES ON ROAD SAFETY*. https://trid.trb.org/view/472973

Hauer, E., & Bamfo, J. (1997). *Two tools for finding what function links the dependent variable to the explanatory variables* ICTCT, Lund, Sweden.

Hauer, E., & Gårder, P. (1986). Research into the validity of the traffic conflict technique. *Accident Analysis & Prevention*, *18*(6), 471-481. https://doi.org/http://dx.doi.org/10.1016/0001-4575(86)90020-5

Hayward, J. C. (1971). *Near misses as a measure of safety at urban intersections* [Master thesis, The Pennsylvania State University].

Hupfer, C. (1997, 5-7 November 1997). *Deceleration to Safety Time (DST) - a useful figure to evaluate traffic safety?* 10th ICTCT workshop, Lund, Sweden.

Hydén, C. (1977). *A traffic conflict technique for determining risk* (LUTVDG/(TVTT-3005)/1-95/(1977) & Bulletin 15 B).

Hydén, C. (1987). *The development of a method for traffic safety evaluation: the Swedish traffic conflict technique* [Doctoral thesis, Lund University].

Ismail, K., Sayed, T., & Saunier, N. (2010). Automated analysis of pedestrian-vehicle conflicts: context for before-and-after studies. *Transportation Research Record*, *2198*, 52-64. https://doi.org/http://doi.org/10.3141/2198-07

Ismail, K., Sayed, T., & Saunier, N. (2011). Methodologies for aggregating indicators of traffic conflict. *Transportation Research Record*, 10-19. https://doi.org/http://doi.org/10.3141/2237-02

Janstrup, K. H., Kaplan, S., Hels, T., Lauritsen, J., & Prato, C. G. (2016). Understanding traffic crash under-reporting: Linking police and medical records to individual and crash characteristics. *Traffic Injury Prevention*, *17*(6), 580-584. https://doi.org/https://doi.org/10.1080/15389588.2015.1128533

Jensen, S. U., & Transportation Research, B. (2008). Bicycle Tracks and Lanes: A Before-and-After Study.

Johnsson, C., Laureshyn, A., Dágostino, C., & De Ceunynck, T. (2020). The ‘safety in density’ effect for cyclists and motor vehicles in Scandinavia: An observational study. *IATSS Research*. https://doi.org/10.1016/j.iatssr.2020.08.003

Johnsson, C., Laureshyn, A., & De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. *Transport Reviews*, 1-21. https://doi.org/10.1080/01441647.2018.1442888

Kamaluddin, N. A., Rahman, M. F. A., & Várhelyi, A. (2019). Matching of police and hospital road crash casualty records – a data-linkage study in Malaysia. *International Journal of Injury Control and Safety Promotion*, *26*(1), 52-59. https://doi.org/https://doi.org/10.1080/17457300.2018.1476385

Kaparias, I., Bell, M. G. H., Greensted, J., Cheng, S., Miri, A., Taylor, C., & Mount, B. (2010). Development and implementation of a vehicle-pedestrian conflict analysis method: Adaptation of a vehicle-vehicle technique. *Transportation Research Record*, 75-82. https://doi.org/http://dx.doi.org/10.3141/2198-09

Kocárková, D. (2012, 29-31 October 2012). *Traffic Conflict Techniques in Czech Republic* SIIV - 5th International Congress - Sustainability of Road Infrastructures, Rome, Italy.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kraay, J. H., van der Horst, A. R. A., & Oppe, S. (2013). Manual conflict observation technique DOCTOR. *Foundation Road safety for all, The Netherlands*.

Kruysse, H. W. (1991). The subjective evaluation of traffic conflicts based on an internal concept of dangerousness. *Accident Analysis & Prevention*, *23*(1), 53-65. https://doi.org/http://dx.doi.org/10.1016/0001-4575(91)90035-4

Kruysse, H. W., & Wijlhuizen, G. J. (1992). Why are experts not better in judging the danger of filmed traffic conflicts? *Accident Analysis & Prevention*, *26*(3), 227-235. https://doi.org/http://dx.doi.org/10.1016/0001-4575(92)90002-Z

Kuang, Y., Qu, X., & Wang, S. (2015). A tree-structured crash surrogate measure for freeways. *Accident Analysis & Prevention*, *77*, 137-148. https://doi.org/http://dx.doi.org/10.1016/j.aap.2015.02.007

Laureshyn, A. (2010). *Application of automated video analysis to road user behaviour* (Publication Number Bulletin 253) [Doctoral thesis, Lund University].

Laureshyn, A., De Ceunynck, T., Karlsson, C., Svensson, Å., & Daniels, S. (2017). In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator. *Accident Analysis & Prevention*, *98*, 46-56. https://doi.org/http://dx.doi.org/10.1016/j.aap.2016.09.026

Laureshyn, A., Johnsson, C., De Ceunynck, T., Svensson, Å., de Goede, M., Saunier, N., Włodarek, P., van der Horst, A. R. A., & Daniels, S. (2016). *Review of current study methods for VRU safety. Appendix 6 – Scoping review: surrogate measures of safety in site-based road traffic observations* (Deliverable 2.1 – part 4).

Laureshyn, A., Svensson, Å., & Hydén, C. (2010). Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation. *Accident Analysis & Prevention*, *42*, 1637-1646. https://doi.org/http://dx.doi.org/10.1016/j.aap.2010.03.021

Laureshyn, A., Svensson, Å., & Hydén, C. (2010). Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation.

Accident Analysis & Prevention, *42*(6), 1637-1646. https://doi.org/http://dx.doi.org/10.1016/j.aap.2010.03.021

Lefèvre, S., Vasquez, D., & Laugier, C. (2014). A survey on motion prediction and risk assessment for intelligent vehicles. *1*(1), 1. https://doi.org/10.1186/s40648-014-0001-z

Lord, D. (1996). Analysis of Pedestrian Conflicts with Left-Turning Traffic. *Transportation Research Record*, *1538*, 61-67. https://doi.org/http://dx.doi.org/10.3141/1538-08

Lord, D., & Mannering, F. (2010). The statistical analysis of crash–frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, *44*(5), 291–305. https://doi.org/https://doi.org/10.1016/j.tra.2010.02.001

Lu, G., Liu, M., Wang, Y., & Yu, G. (2012). Quantifying the Severity of Traffic Conflict by Assuming Moving Elements as Rectangles at Intersection. *Procedia - Social and Behavioral Sciences*, *43*, 255-264. https://doi.org/http://dx.doi.org/10.1016/j.sbspro.2012.04.098

Lägnert, J. (2019). *An Extreme Value Approach to Road Safety Analysis* Lund University].

Madsen, T., Bahnsen, C., Jensen, M., Lahrmann, H., & Moeslund, T. (2016). *Watchdog System* (Deliverable 4.1).

Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic methods in accident research*, *1*, 1-22. https://doi.org/https://doi.org/10.1016/j.amar.2013.09.001

Migletz, D. J., Glauz, W. D., & Bauer, K. M. (1985). *Relationships between traffic conflicts and accidents* (FHWA/RD-84/042).

Minderhoud, M. M., & Bovy, P. H. L. (2001). Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention*, *33*(1), 89-97. https://doi.org/http://dx.doi.org/10.1016/S0001-4575(00)00019-1

Mohamed, M. G., & Saunier, N. (2013). Motion Prediction Methods for Surrogate Safety Analysis. *Transportation Research Record*, *2386*, 168-178. https://doi.org/http://dx.doi.org/10.3141/2386-19

Morse, P., St-Aubin, P., Miranda-Moreno, L., & Saunier, N. (2016, 10-14 January 2016). *Transferability Study of Video Tracking Optimization for Traffic Data Collection and Analysis* 95th TRB Annual Meeting, Washington DC, USA.

Muhlrad, N., & Dupre, G. (1984). *The French conflict technique* NATO advanced research workshop on international calibration study of traffic conflict techniques, Copenhagen, Denmark.

Mulvaney, C. A., Smith, S., Watson, M. C., Parkin, J., Coupland, C., Miller, P., Kendrick, D., & McClintock, H. (2015). Cycling infrastructure for reducing cycling injuries in cyclists. *Cochrane Database of Systematic Reviews*(12). https://doi.org/10.1002/14651858.CD010415.pub2

Ogawa, K. (2007). An Analysis of Traffic Conflict Phenomenon of Bicycles Using Space Occupancy Index. *Journal of the Eastern Asia Society for Transportation Studies*, *7*, 1820-1827.

Olszewski, P., Osińska, B., Szagała, P., Skoczyński, P., & Zielińska, A. (2016). Problems with assessing safety of vulnerable road users based on traffic accident data. *Archives of Civil Engineering*, *LXII*(4). https://doi.org/https://doi.org/10.1515/ace-2015-0113

Parker, M. R., & Zegeer, C. V. (1989). *Traffic Conflict Techniques for Safety and Operation* (FHWA-IP-88-027).

Prati, G., Marín Puchades, V., De Angelis, M., Fraboni, F., & Pietrantoni, L. (2018). Factors contributing to bicycle–motorised vehicle collisions: a systematic literature review. *Transport Reviews*, *38*(2), 184-208. https://doi.org/10.1080/01441647.2017.1314391

Risser, R., & Schutzenhofer, A. (1984). *Application of traffic conflict technique in Austria* NATO advanced research workshop on international calibration study of traffic conflict techniques, Copenhagen, Denmark.

Sacchi, & Sayed. (2016). Conflict-Based Safety Performance Functions for Predicting Traffic Collisions by Type. *Transportation Research Record: Journal of the Transportation Research Board*, *2583*, 50-55. https://doi.org/10.3141/2583-07

Sacchi, Sayed, T., & de Leur, P. (2013). A comparison of collision-based and conflict-based safety evaluations: the case of right-turn smart channels. *Accident Analysis & Prevention*, *59*, 260-266. https://doi.org/http://doi.org/10.1016/j.aap.2013.06.002

Saunier, N., Sayed, T., & Ismail, K. (2010). Large scale automated analysis of vehicle interactions and collisions. *Transportation Research Record*, *2147*, 42-50. https://doi.org/http://dx.doi.org/10.3141/2147-06

Sayed, T., & Zein, S. (1999). Traffic conflict standards for intersections. *Transportation Planning & Technology*, *22*, 309-323. https://doi.org/http://dx.doi.org/10.1016/j.ssci.2012.04.015

Shadish, W., Cook, T., & Campbell, D. (2002). Statistical Conclusion Validity and Internal Validity. In W. Shadish, T. Cook, & D. Campbell (Eds.), *Experimental and Quasi Experimental Designs for Generilized Causal Inference* (pp. 33-63). Houghton Mifflin.

Shbeeb, L. (2000). *Development of a traffic conflicts technique for different environments - a comparative study of pedestrian conflicts in Sweden and Jordan* [Doctoral thesis, University of Lund, Lund Institute of Technology].

Shelby, S. G. (2011, 23-27 January 2011). *Delta-V as a measure of traffic conflict severity* 90th TRB Annual Meeting, Washington DC, USA.

Shinar, D. (1984). The traffic conflict technique: a subjective vs. objective approach. *Journal of safety research*, *15*, 153-157. https://doi.org/http://dx.doi.org/10.1016/0022-4375(84)90046-X

Songchitruksa, P., & Tarko, A. (2006). The extreme value theory approach to safety estimation. *Accident Analysis and Prevention*, *38*, 811-822. https://doi.org/http://dx.doi.org/10.1016/j.aap.2006.02.003

St-Aubin, P., Saunier, N., & Miranda-Moreno, L. F. (2014, 12-16 January 2014). *Road user collision prediction using motion patterns applied to surrogate safety analysis* 93rd TRB Annual Meeting, Washington DC, USA.

Summala, H., Pasanen, E., Räsänen, M., & Sievänen, J. (1996). Bicycle accidents and drivers' visual search at left and right turns. *Accident Analysis and Prevention*, *28*, 147-153.

Svensson. (1992). *Vidareutveckling och validering av den svenska konflikttekniken* ((unpublished)).

Svensson. (1998). *A method for analysing the traffic process in a safety perspective* [Doctoral thesis, University of Lund, Lund Institute of Technology].

Svensson, Å. (1992). *Vidareutveckling och validering av den svenska konflikttekniken*.

Svensson, Å. (1998). *A method for analysing the traffic process in a safety perspective* [Doctoral thesis, University of Lund, Lund Institute of Technology].

T-Analyst. (2019). *T-Analyst. Semi-automated video processing*. Retrieved 15 Jan from https://bitbucket.org/TrafficAndRoads/tanalyst/wiki/Manual

Tageldin, A., & Sayed, T. (2016). Developing evasive action-based indicators for identifying pedestrian conflicts in less organized traffic environments. *Journal of Advanced Transportation*, *50*, 1193-1208. https://doi.org/http://doi.org/10.1002/atr.1397

Tageldin, A., Sayed, T., & Wang, X. (2015, 11-15 January 2015). *Can Time Proximity Measures Be Used as Safety Indicators in All Driving Cultures? A Case Study of Motorcycle Safety in China* 94th TRB Annual Meeting, Washington DC, US. http://docs.trb.org/prp/15-4371.pdf

http://trid.trb.org/view/1338588

Tarko, A. P., Romero, M. A., Bandaru, V. K., Ariyur, K. B., & Lizarazo, C. G. (2017, 8-12 January 2017). *Feasibility of Tracking Vehicles at Intersections with a Low-end LiDAR* 96th TRB Annual Meeting,

Thomas, B., & DeRobertis, M. (2013). The safety of urban cycle tracks: A review of the literature. *Accident Analysis & Prevention*, *52*, 219-227. https://doi.org/http://dx.doi.org/10.1016/j.aap.2012.12.017

Thompson, E., Clem, N., Peter, D. A., Bryan, J., Peterson, B. I., & Holbrook, D. (2015). Parallel cuda implementation of conflict detection for application to airspace deconfliction. *Journal of Supercomputing*, *71*(10), 3787-3810. https://doi.org/10.1007/s11227-015-1467-z

Thompson, J., Savino, G., & Stevenson, M. (2015). Reconsidering the Safety in Numbers Effect for Vulnerable Road Users: An Application of Agent-Based Modeling. *Traffic Injury Prevention*, *16*(2), 147-153. https://doi.org/10.1080/15389588.2014.914626

Thompson, J., Savino, G., & Stevenson, M. (2016). A model of behavioural adaptation as a contributor to the safety-in-numbers effect for cyclists. *Transportation Research Part A: Policy and Practice*, *85*, 65-75. https://doi.org/https://doi.org/10.1016/j.tra.2015.12.004

Thompson, J., Wijnands, J. S., Savino, G., Lawrence, B., & Stevenson, M. (2017). Estimating the safety benefit of separated cycling infrastructure adjusted for behavioral adaptation among drivers; an application of agent-based modelling. *Transportation Research Part F: Traffic Psychology and Behaviour*, *49*, 18-28. https://doi.org/https://doi.org/10.1016/j.trf.2017.05.006

Thompson, J. H., Wijnands, J. S., Mavoa, S., Scully, K., & Stevenson, M. R. (2018). Evidence for the 'safety in density' effect for cyclists: validation of agent-based modelling results. *Injury Prevention*, injuryprev-2018-042763. https://doi.org/10.1136/injuryprev-2018-042763

Tsai, R. Y. (1986). *An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision* IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL.

van der Horst, A. R. A. (1984). *The ICTCT calibration study at Malmö: a quantitative analysis of video recordings* (IZF 1984-37).

van der Horst, A. R. A. (1991). Time-to-Collision as a cue for decision making in braking. In A. G. Gale (Ed.), *Vision in vehicles - III* (pp. 19-26). Elsevier Science Publishers B.V.

van der Horst, A. R. A., Rook, A. M., Amerongen, P. J. M. v., & Bakker, P. J. (2007). Video-recorded accidents, conflicts and road user behaviour: Integral Approach Analysis of Traffic Accidents (IAAV). *TNO Defence, Security and Safety*.

van der Horst, A. R. A., Thierry, M. C., Vet, J. M., & Rahman, A. F. (2016). An evaluation of speed management measures in Bangladesh based upon alternative accident recording, speed measurements, and DOCTOR traffic conflict observations. *Transportation Research Part F*(in press). https://doi.org/http://dx.doi.org/10.1016/j.trf.2016.05.006

Wang, C., & Stamatiadis, N. (2014). Evaluation of a simulation-based surrogate safety metric. *Accident Analysis & Prevention*, *71*, 82-92. https://doi.org/http://dx.doi.org/10.1016/j.aap.2014.05.004

Várhelyi, A. (1998). Driver's speed behaviour at a zebra crossing: a case study. *Accident Analysis & Prevention*, *30*(6), 731-743. https://doi.org/http://dx.doi.org/10.1016/S0001-4575(98)00026-8

Yannis, G., Papadimitriou, E., Chaziris, A., & Broughton, J. (2014). Modeling road accident injury under-reporting in Europe. *European Transport Research Review*, *6*(4), 425-438. https://doi.org/https://doi.org/10.1007/s12544-014-0142-4

Zheng, L., Ismail, K., & Meng, X. (2014a). Freeway safety estimation using extreme value theory approaches: a comparative study. *Accident Analysis & Prevention*, *62*, 32-41. https://doi.org/http://dx.doi.org/10.1016/j.aap.2013.09.006

Zheng, L., Ismail, K., & Meng, X. (2014b). Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes. *Safety Science*, *64*, 155-162. https://doi.org/http://dx.doi.org/10.1016/j.ssci.2013.12.003

Zheng, L., Ismail, K., & Meng, X. (2014c). Traffic conflict techniques for road safety analysis: open questions and some insights. *Canadian Journal of Civil Engineering*, *41*(7), 633-641. https://doi.org/http://doi.org/10.1139/cjce-2013-0558

13. List of included papers

Paper 1: In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators, Carl Johnsson, Aliaksei Laureshyn & Tim De Ceunynck, Published in Transport Reviews (2018), Volume 38, Pages 765-785

Paper 2: Validation of surrogate measures of safety with a focus on cyclist-motor vehicle interactions, Carl Johnsson, Aliaksei Laureshyn & Carmelo D'agostino Submitted to Accident Analysis and Prevention

Paper 3: A relative approach in validation of surrogate measures of safety Carl Johnsson, Aliaksei Laureshyn & Carmelo D'agostino, Submitted to Accident Analysis and Prevention

Paper 4: The *safety in density* effect for cyclists and motor vehicles in Scandinavia: An observational study, Carl Johnsson, Aliaksei Laureshyn, Carmelo D'agostino & Tim De Ceunynck, Article in press in IATSS Research

Paper 5: A general method for identifying evasive actions and calculating time to collision from trajectory data, Carl Johnsson, Submitted to Accident Analysis and Prevention

Declaration of contribution

In paper 1, Johnsson did approximately half of the data collection and was the main contributor to the result and the text. For paper 2 and paper 3, Johnsson was involved with overseeing and planning the data collection. The result and analysis were mainly made by Johnsson and D'agostino. In both cases, the main text was written by Johnsson with help from the other authors. For paper 4, Johnsson was partly responsible for the data collection. The result and analysis were made by Johnsson with help from D'agostino. The text was mainly written by Johnsson with help from the other authors. The fifth paper was made solely by Johnsson using data described in paper 2.